

529, 263

10/529263

(12)特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関  
国際事務局



(43) 国際公開日  
2004 年 4 月 8 日 (08.04.2004)

PCT

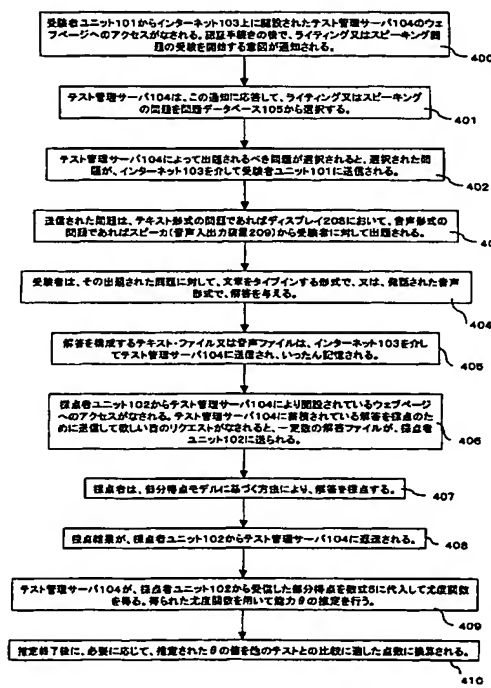
(10) 国際公開番号  
WO 2004/029906 A1

- (51) 国際特許分類: G09B 19/00, G06F 17/18, 17/60
- (21) 国際出願番号: PCT/JP2003/012252
- (22) 国際出願日: 2003 年 9 月 25 日 (25.09.2003)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (30) 優先権データ:  
特願2002-279513 2002 年 9 月 25 日 (25.09.2002) JP
- (71) 出願人 (米国を除く全ての指定国について): 株式会社ベネッセコーポレーション (BENESSE CORPORATION) [JP/JP]; 〒700-8686 岡山県岡山市南方 3-7-17 Okayama (JP).
- (71) 出願人 および
- (72) 発明者: 藤森進 (FUJIMORI, Susumu) [JP/JP]; 〒183-0042 東京都府中市武蔵台 2-3 6-1 O Tokyo (JP).
- (72) 発明者; および
- (75) 発明者/出願人 (米国についてのみ): 山下 仁司 (YAMASHITA, Hitoshi) [JP/JP]; 〒206-8686 東京都多摩市落合 1 丁目 3 4 番地 株式会社ベネッセコーポレーション内 Tokyo (JP).
- (74) 代理人: 社本 一夫, 外 (SHAMOTO, Ichio et al.); 〒100-0004 東京都千代田区大手町二丁目 2 番 1 号 新大手町ビル 206 区 ユアサハラ法律特許事務所 Tokyo (JP).
- (81) 指定国 (国内): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

[続葉有]

(54) Title: TEST SYSTEM AND CONTROL METHOD THEREOF

(54) 発明の名称: テスト・システム及びその制御方法



400...ACCESS IS MADE FROM EXAMINEE UNIT (101) TO WEB PAGE OF TEST MANAGEMENT SERVER (104) ESTABLISHED ON INTERNET (103). AFTER AUTHENTICATION PROCESS, INTENTION TO TAKE EXAMINATION ON WRITING OR SPEAKING IS NOTIFIED.

401...IN RESPONSE TO THIS NOTIFICATION, THE TEST MANAGEMENT SERVER (104) SELECTS A WRITING OR SPEAKING QUESTION FROM A QUESTION DATABASE (105).

402...WHEN A QUESTION TO BE PRESENTED IS SELECTED BY THE TEST MANAGEMENT SERVER (104), THE SELECTED QUESTION IS TRANSMITTED VIA THE INTERNET (103) TO THE EXAMINEE UNIT (101).

403...THE TRANSMITTED QUESTION IS PRESENTED TO THE EXAMINEE VIA DISPLAY (208) IF THE QUESTION IS IN A TEXT FORM OR VIA LOUDSPEAKER (AUDIO I/O DEVICE (209)) IF THE QUESTION IS AN AUDIO FORM.

404...EXAMINEE ANSWERS THE QUESTION BY TYPING IN A SENTENCE OR SPEAKING IN AN AUDIO FORM.

405...TEXT FILE OR AUDIO FILE CONSTITUTING THE ANSWER IS TRANSMITTED VIA THE INTERNET (103) TO THE TEST MANAGEMENT SERVER (104) AND STORED.

406...MARKER UNIT (102) ACCESSES THE WEB PAGE CREATED BY THE TEST MANAGEMENT SERVER (104). WHEN A REQUEST IS MADE TO TRANSMIT ANSWERS ACCUMULATED IN THE TEST MANAGEMENT SERVER (104) FOR MAKING, A PREDETERMINED NUMBER OF ANSWER FILES ARE TRANSMITTED TO THE MARKER UNIT (102).

407...MARKER MARKS THE ANSWER ACCORDING TO A METHOD BASED ON PARTIAL POINT MODEL.

408...THE MARKING RESULT IS RETURNED FROM THE MARKER UNIT (102) TO THE TEST MANAGEMENT SERVER (104).

409...THE TEST MANAGEMENT SERVER (104) SUBSTITUTES THE PARTIAL POINT RECEIVED FROM THE MARKER UNIT (102) IN EQUATION 5 TO OBTAIN A LIKELIHOOD FUNCTION. THE OBTAINED LIKELIHOOD FUNCTION IS USED TO ESTIMATE ABILITY  $\theta$ .

410...AFTER THE ESTIMATION, IF NECESSARY, THE VALUE OF THE ESTIMATED  $\theta$  IS CONVERTED INTO A POINT APPROPRIATE FOR COMPARISON WITH OTHER TESTS.

(57) Abstract: There is provided a test system capable of giving a partial point in addition to the binary evaluation of correct/incorrect in an English achievement test or the like. The test system includes computers (101, 102) having I/O devices (207, 208) and a network (103) such as the Internet. Estimation of an item parameter and ability utilizes a partial point model obtained by modifying the conventional item reaction theory.

[続葉有]

WO 2004/029906 A1



(84) 指定国 (広域): ARIPO 特許 (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), ユーラシア特許 (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), ヨーロッパ特許 (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI 特許 (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

添付公開書類:

— 国際調査報告書

2文字コード及び他の略語については、定期発行される各PCTガゼットの巻頭に掲載されている「コードと略語のガイダンスノート」を参照。

---

In this partial point model, a partial point is assumed to be composed of a correct/incorrect average of a plurality of items having the same parameter. By using this test system, it becomes easier than in the conventional system to control a test of a question form which is difficult to be judged only by correct/incorrect and requires a partial point such as writing and speaking in a foreign language test.

(57) 要約: 本発明によると、英語能力テストなどにおいて正誤の2値的な評価だけでなく部分得点を与える採点を可能にするテスト・システムが提供される。このテスト・システムでは、入出力装置(207、208)を含むコンピュータ(101、102)とインターネットなどのネットワーク(103)とが利用され、項目パラメータ及び能力の推定に従来の項目反応理論を修正した部分得点モデルが用いられる。この部分得点モデルでは、部分得点を、同一のパラメータを有する複数の項目の正誤の平均として構成されるものと想定している。本発明のテスト・システムを用いることにより、外国語テストにおけるライティングやスピーキングのように正誤だけでは判定困難であって部分得点を要するような出題形式のテストの制御が、従来よりも容易になる。

## 明細書

### テスト・システム及びその制御方法

#### 5 技術分野

本発明は、テスト・システム及びその制御方法に関する。更に詳しくは、本発明は、従来型の項目反応理論を修正した部分得点モデルに依拠し、採点の際に単なる正誤の2値だけでなく複数の段階を有する部分得点としての評価も許容するテストを従来よりも簡略化された態様で設計、実施及び評価するテスト・システム及びその制御方法に関する。

#### 背景技術

今日、テストを設計しその結果を処理する際に用いられる理論として、古典的テスト理論と項目反応理論との2つが広く知られている。これらのテスト理論に  
15 関しては、例えば、渡部洋編著『心理統計の技法』（福村出版、2002年）の6章及び7章にそれぞれ一般的な解説が与えられている。なお、テスト理論におけるテストとは、学力テストだけでなく、心理学における性格検査や臨床検査なども含むのであるが、この出願では、そのような広い応用分野を視野に入れた抽象的な説明を行うのではなく、理解を容易にするために、特に外国語試験などの  
20 学力テストを念頭において具体的な説明を試みたい。なお、項目反応理論における「項目」とは、学力テストの場合であれば問題を意味する。従って、この出願で具体例に即した説明する際には、「項目」ではなく「問題」、「被験者」ではなく「受験者」など、学力テストにおける一般的な用語を用いる。しかし、本発明は、一般のテスト理論に関する汎用的なものであって、学力テストだけに限定  
25 されることはなく、テスト理論が一般的に適用される上記の分野にも適用が可能である。

古典的テスト理論は、現在広く用いられているが、得られる結論が受験者の特性の統計的な分布に依存するという特徴を有する。従って、古典的テスト理論に依拠している場合には、特性が異なる受験者の集団に対して実施された複数のテ

ストで得られた平均点や偏差値を比較することは、適切でない。

古典的テスト理論の短所を克服する理論として項目反応理論があり、現在では、この項目反応理論に基づいて設計され結果が処理される学力テストが多く存在する。例えば、最近出版された項目応答理論の教科書である豊田秀樹著『項目反応理論入門編』（朝倉書店、2002年）の第16頁には次のような記述がある。

「有名な語学試験であるTOEFL…は1年間に何度も実施される。しかも世界中で実施される。同一の被験者が再受験する可能性があるから同じ項目の集まりである同一のテストは2度と使用できない。このため平均点や通過率は、テストごとにそれぞれ異なる。地域によって英語力にも差があるから特性値の分布も異なる。したがって、偏差値や正規得点や経験分布関数を使用して、受験結果を異なるテスト間で比較することは不可能である。…TOEFLの点（たとえば、500点、650点など）は、どこで受験したか、いつ受験したか、どの問題を解いたかに関係なく留学の可否の判断に利用される。つまり異質な受験者が、異なる項目を、異なる日時に、異なる場所で受験したにも関わらず、被験者は、統一された処遇を受けることができる。この処遇を可能にするテストを継続的に、前向きに運用するシステムを構築する数理モデルが、…項目反応モデルである。」

また、特開2002-006734号公報には、出題が予定される試験問題を予め樹状に配置しておき、受験者による解答の正否に応じて樹状に配置された経路に沿って問題を順に出題し、単に正解の数だけでなく、どのような経路で最終地点まで到達したかという途中経過を考慮して受験者の能力を推定するテストの方法及びシステムが開示されている。この公開特許公報にも、項目反応理論への言及が見られる。

しかし、従来型の理論に基づく項目反応モデルでは、2値によって採点が可能な○×式の解答以外の、例えば、1と0との間の部分点を許容せざるを得ないような形式の問題の処理が容易ではなかった。段階反応モデルなど、多値の結果も扱えるように構築された方法も古くから存在してはいるが、従来の方法は後述する短所を有している。従って、部分得点を許容しなければ的確な採点が困難な英作文やスピーチなどは、コンピュータ適応型テスト（CAT）に含まれないことが通常である。例えば、上述の特開2002-006734号公報に開示されて

いる発明では、項目反応理論の利用が明言され、出題されるべき問題が予め樹状に配置されているが、受験者がある問題に正解したら右下、誤答したら左下に配置された問題に移動するというように、解答が正誤のいずれかの2値であることが予定されている。

5

### 発明の開示

本発明は、このような従来型の2値の解答を予想するCATとは異なり、部分得点を与えるような採点を許容するテスト・システムであって、しかも、従来の段階反応モデル等の複雑なモデルよりもはるかに容易に部分得点の処理を可能にするテスト・システム及びその制御方法を提供することを目的とする。

10

本発明によると、入力装置と出力装置とを備えた第1のコンピュータと、インターネットを含むネットワークを介して前記第1のコンピュータと接続されており入力装置と出力装置とを備えた第2のコンピュータと、前記ネットワークを介して前記第1及び第2のコンピュータと接続されたテスト管理サーバと、前記テスト管理サーバからアクセス可能であって難易度と識別力とを含む項目パラメータが予め推定されている複数の問題が記憶された問題データベースとによって構成され、1人の受験者に $n$ 問の問題を出題し、出題された $n$ 問の問題に対する前記受験者の反応から前記受験者の能力 $\theta$ を推定するテスト・システムが提供される。前記テスト管理サーバは、(1)前記第1のコンピュータから送信されたり

クエストに回答して、 $1 \leq j \leq n$ である問題 $j$ に対し満点を1として $0 \leq r_j \leq 1$ である部分得点 $r_j$ を許容する態様で採点されうる $n$ 問の問題を前記問題データベースから選択し前記第1のコンピュータに送信する手段と、(2)前記問題データベースから選択され前記第1のコンピュータに送信された問題に対して前記第1のコンピュータから返送された解答を記憶する解答記憶手段と、(3)前記第2のコンピュータから送信されたりクエストに回答して、前記解答記憶手段に記憶されている解答を読み出し前記第2のコンピュータに送信する手段と、(4)前記第2のコンピュータに送信された解答に対して与えられた部分得点 $r_j$ を前記第2のコンピュータから受信し記憶する部分得点記憶手段と、(5)前記部分得点記憶手段に記憶されている部分得点 $r_j$ と前記問題データベースに記

15

20

25

憶されている問題  $j$  の項目パラメータとを用いて、当該部分得点  $r_j$  を獲得した受験者の能力  $\theta$  を推定する能力推定手段とを備えている。そして、(5) の前記能力推定手段においては、 $P_j(\theta)$  は、部分得点  $r_j$  を問題  $j$  に固有であって前記受験者が正答 1 又は誤答 0 のいずれか一方の潜在的反応を取り得る潜在的問題を  $s_j$  回繰り返し実施したときに前記受験者が潜在的に取りうる正誤反応の平均であると想定した場合に、前記受験者が前記潜在的問題に正答する確率であり、

【数 1】

$$P_j(\theta) = \frac{1}{1 + \exp(-Da_j(\theta - b_j))}$$

と表現され、この数式 1 における  $a_j$  及び  $b_j$  はそれぞれが前記問題データベースに記憶されている問題が有する固有の特性である識別力と困難度とであり、 $D$  は 1.7 という定数であり、 $Q_j(\theta)$  は  $1 - P_j(\theta)$  である場合に、

【数 2】

$$\ell_{part}(\theta) = \sum_{j=1}^n s_j (r_j \ln(P_j(\theta)) + (1 - r_j) \ln(Q_j(\theta)))$$

によって表される対数尤度  $\ell_{part}(\theta)$  を用いて受験者の能力  $\theta$  が推定される。ただし、上記の数式 1 として表現されている  $P_j(\theta)$  の関数形は単なる例示であり、 $P_j(\theta)$  がこの表現形式に限定されることは必要なく、様々な形式でありうる。

ここで注意すべきは、観測可能なものは、受験者が問題  $j$  に対して獲得する 0 から 1 までの間の値を取りうる部分得点  $r_j$  であり、潜在的問題に対する受験者の潜在的な反応は観測できないという点である。

更に、本発明によるテスト・システムによれば、実際に観測される問題  $j$  に対する部分得点  $r_j$  が顕在的な複数の正誤問題の平均によって構成されている場合にも、これらの正誤問題の正答確率の平均を数式 1 で表現し、数式 2 を用いて受験者の能力  $\theta$  を推定することが可能である。

更に、本発明によるテスト・システムによれば、テストが実施された集団の能力分布を仮定した上で、 $s_j$  回の正誤の和である 2 項分布と仮定された能力分布との積を能力の次元で積分して部分得点の理論的な分布関数を求め、求められた理論的な分布関数と実際のデータの部分得点の経験的な分布関数が最もよく一致す

るように、潜在的な問題の繰り返し回数である  $s_j$  を推定することも可能である。

更に、本発明によるテスト・システムにおいて、前記第1及び第2のコンピュータにおける前記出力装置と前記入力装置とがそれぞれ音声出力装置と音声入力装置とを含む場合には、前記テストサーバに送信され記憶される解答として音声  
5 データを含ませることも可能である。この場合には、第1のコンピュータにおいてリスニングの問題を出題したり、受験者に実際に発話させた内容を採点対象とするスピーキングの問題も出題可能となる。

なお、本発明は、上述のテスト・システムを制御する方法としても実現することができる。更に、本発明は、そのようなテスト・システム制御方法を実装する  
10 コンピュータ・プログラムが記憶されているコンピュータ可読な記憶媒体自体として存在することもありうる。更には、そのようなテスト・システム制御方法を実行するコンピュータ・プログラム自体としても存在しうる。

#### 図面の簡単な説明

15 図1は、本発明によるテスト・システムの一例の概要である。

図2は、本発明によるテスト・システムを構成する受験者ユニットの概要である。

図3は、本発明によるテスト・システムを構成する採点者ユニットの概要である。

20 図4は、本発明によるシステムを用いたテスト実施の概要を示す流れ図であり、特に、本発明が依拠する部分得点に関するライティング及びスピーキングに関する受験及び採点のプロセスが示されている。

図5は、本発明によるテスト・システムを用いた能力推定の有効性を確認するために実施された被験者12名によるスコア安定性確認調査の結果を示すグラフ  
25 である。

図6は、図6aから図6gで構成され、それぞれが、図5のスコア安定性確認調査における被験者12名のスコアを示すグラフである。

図7は、図7aから図7dで構成され、繰り返し数  $s_j$  を推定する際に、真の  $s = 5, 10, 20, 40$  の場合に、推定された経験分布と理論分布との分布関

数の差の最大値（コルゴモロフ・スミルノフ検定の統計量）を繰り返し数 3 ～ 10 についてプロットしたものである。

図 8 は、図 8 a 及び図 8 b で構成され、それぞれが、英語能力テストでの繰り返し回数  $s_j$  の推定への適用例である。

5

#### 発明を実施するための最良の形態

本発明の基礎として用いる部分得点モデルについて述べる前に、一般的な 2 値の項目反応モデルについて簡単に説明する。2 値とは、解答が正答又は誤答の 2 つの値だけを取るという意味である。このモデルでは、受験者が問題に正答する確率を、受験者の能力を表すパラメータと問題の特徴付けるパラメータとを用いて表現する。例えば、本発明では、各問題を 2 つのパラメータ（識別力  $a$  及び困難度  $b$ ）によって特徴付ける 2 母数（パラメータ）ロジスティック・モデルを用いるのであるが、この場合、能力  $\theta$  を有する受験者  $i$  が問題  $j$  に正答する確率は次のように書くことができる。

15 【数 3】

$$P_j(x_{ij} = 1 | \theta_i) = \frac{1}{1 + \exp(-Da_j(\theta_i - b_j))}$$

ここで、 $x$  は、受験者  $i$  が問題  $j$  に正答するならば 1 であり、誤答するならば 0 であるダミー変数である。また、 $D$  は定数である。この場合に、 $n$  問の問題を終了した時点での受験者の能力  $\theta$  の尤度  $L_B(\theta)$  は、次のように書くことができる。

20

【数 4】

$$L_B(\theta) = \prod_j P_j^{x_j}(\theta) Q_j^{1-x_j}(\theta)$$

ここで、 $P(\theta)$  は数式 3 の右辺の正答確率であり、 $Q(\theta)$  は誤答確率つまり  $1 - P(\theta)$  である。項目応答理論では、数式 4 の尤度  $L_B(\theta)$  の最大値を与える  $\theta$  の値を受験者の能力パラメータの推定値とする最尤推定法が知られ、広く用いられている。ただし、数式 4 の右辺は積の形式で書かれており最大値を求めるのが容易でないので、和の形式に直して考察するため、両辺の自然対数を取っ

25



た対数尤度  $\ln(L_B(\theta))$  の最大値を求めるのが一般的である。これは、自然対数は単調増加関数であって、尤度  $L_B(\theta)$  の最大値を与える  $\theta$  とその自然対数  $\ln(L_B(\theta))$  の最大値を与える  $\theta$  とが一致するからである。

- 5 以上が、従来から公知である一般的な2母数ロジスティック・モデルであるが、本発明において用いる部分得点モデルでは、問題への反応（解答）の評価が正誤という2値にとどまらず、部分点として評価されることが可能となる。具体的には、数式4のダミー変数  $x_j$  が1及び0の2値だけでなく、0から1までの3つ以上の複数の値（例えば、0、0.2、0.4、0.6、0.8、1の6つの値）を取ることを許容する。受験者  $i$  の問題  $j$  に対する部分得点を  $r_{ij}$  とすると、部分得点に対応する尤度は、次のように表すことができる。

【数5】

$$L_{part}(\theta_i) = \prod_j^n P_j^{r_{ij}}(\theta_i) Q_j^{1-r_{ij}}(\theta_i)$$

- 15 具体的な外国語試験、例えば英語の試験を考えると、正誤問題や多肢選択式問題であれば、正答及び誤答の2値として処理できる。しかし、英作文や英語での発話（質問に対する口頭での応答）は単純な正答又は誤答として評価するのは困難であり、部分点を与えることによって評価することが必要となる。従って、従来型の2値の項目応答モデルを用いることはできない。しかし、数式5によれば、任意の形式を有するテストの採点結果を部分得点に変換して分析用のデータとすることができる。

- 20 数式5の意味を解釈するために、同一の問題パラメータを有する  $s_j$  個の問題が同じ受験者に向けて出題されたとする。学力テストの場合には、パラメータは同じであるが内容としては別の問題を提示すると想定するのが無理のない想定である。性格検査などに用いられる質問紙では、内容的にも同一の問題を提示するケースも考えられるが、これは項目反応理論の前提である局所独立の仮定に反する可能性も生じるので、ここでは、一応同一パラメータであるが内容の異なる質問項目を提示すると想定する。

一般の項目反応モデルでは、このような想定条件下の尤度は、数式4から次のように表すことができる。

## 【数 6】

$$L_B(\theta_i) = \prod_j^n \prod_k^{s_j} P_j^{x_{ijk}}(\theta_i) Q_j^{1-x_{ijk}}(\theta_i) = \prod_j^n P_j^{\sum_k^{s_j} x_{ijk}}(\theta_i) Q_j^{\sum_k^{s_j} (1-x_{ijk})}(\theta_i)$$

すべての問題に関して繰り返しの解答  $s_j$  が等しい ( $=s$ ) と仮定した場合、この数式 6 の  $s$  乗根をとると、次の数式が得られる。

## 5 【数 7】

$$L_B^*(\theta_i) = \sqrt[s]{L_B(\theta_i)} = \sqrt[s]{\prod_j^n P_j^{\sum_k^{s_j} x_{ijk}}(\theta_i) Q_j^{\sum_k^{s_j} (1-x_{ijk})}(\theta_i)} = \prod_j^n P_j^{\frac{\sum_k^{s_j} x_{ijk}}{s}}(\theta_i) Q_j^{1-\frac{\sum_k^{s_j} x_{ijk}}{s}}(\theta_i)$$

$L_B(\theta)$  と  $L_B^*(\theta)$  との最大値を与える  $\theta$  は同一であるので、最尤推定値は、数式 6 及び数式 7 で同じである。数式 7 において

## 【数 8】

$$\frac{\sum_k^{s_j} x_{ijk}}{s} = r_j$$

10

とおけば、 $L_{part}(\theta)$  と  $L_B^*(\theta)$  とは形式上同じとなる。つまり、パラメータが同じである問題の繰り返しによる正誤の平均によって部分得点が表現される場合には、本発明が依拠する部分得点モデル  $L_{part}(\theta)$  の解と一般の項目反応理論による  $L_B(\theta)$  の解とは、と  $L_B^*(\theta)$  を通じて一致する。以上により、同一パラメータの問題を複数回実施し、その正誤の合計又は平均により成績をつけるようなタイプの部分得点に関しては、当該項目のすべての実施結果に関してその平均を取ることににより 0 から 1 までの間の部分得点を作ることになれば、部分得点モデル  $L_{part}(\theta)$  によって処理しても、従来の 2 値データによる処理と同一の結果となることがわかる。数式 8 から、提示の回数  $s_j$  を増加させれば、実質的に 0 から 1 までの任意の部分得点を表現できることが導かれる。なお、数式 6 と数式 7 との間で  $s$  乗根分の相違があると実際には不便なので、次のように、数式 5 を  $s$  乗して、その自然対数を部分得点化の対数尤度とするのが好ましい。

15

20

## 【数 9】

$$\ell_{part} = \sum_{j=1}^n s \{ r_j \ln(P_j(\theta)) + (1-r_j) \ln(Q_j(\theta)) \}$$

さて、実際の学力テストでは、相互に関連する複数の問題が出題されることがある。外国語テストの例では、一定の長さの文章に関して複数の問題が出題される場合などがこれに該当する。しかし、このような問題群の存在は、項目反応理論の局所独立の仮定に反する可能性がある。

因子分析的に考えてみると、仮にテストの得点  $z$  が、テスト問題の全体で測定する因子  $f_1$  と、ある問題群  $J$  に共通する因子  $f_2$  と、各問題に独自の因子  $u$  から構成されるとすれば、受験者  $i$  による問題  $j$  の得点は次のように表現することができる。

10 【数 10】

$$z_{ij} = f_{i1}a_{1j} + f_{i2}a_{2j} + u_j$$

ここで、 $a$  は因子負荷である。問題  $j$  が問題群  $J$  に属する場合には  $a_{2j}$  はゼロでなく、問題  $j$  が問題群  $J$  に属する場合には  $a_{2j} = 0$  である。数式 10 において問題群  $J$  についての和を取ると、次のようになる。

15 【数 11】

$$\sum_j z_{ij} = \sum_j f_{i1}a_{1j} + \sum_j f_{i2}a_{2j} + \sum_j u_j$$

右辺第 2 項は問題群  $J$  に共通する因子  $f_2$  に関する和であるが、因子の直交性の仮定からは、他の問題との相関を持たず独自性部分と見ることができる。項目反応理論でも項目に固有な成分を想定しているわけであるから、右辺第 2 項はこれに該当する。すなわち、局所独立の仮定に触れずに済むことになる。因子  $f_1$  を項目反応理論の能力パラメータ  $\theta$  と仮定すれば、テストで相互に関連する問題群の和を取って部分得点として処理することには不都合はない。局所従属の関係にある問題に関して本発明が提唱するような処理を行うことは、項目反応理論の仮定からはむしろ望ましいとさえいえる。

25 もし完全な同一パラメータの項目でなく類似パラメータの項目の繰り返しを想定する場合は、一般の 2 値の尤度は項目  $j$  の繰り返し提示の正誤パターンによっ

て異なり、次のようになる。

【数 1 2】

$$L_B(\theta_i) = \prod_j^n \prod_k^{s_j} P_{jk}^{x_{jk}}(\theta) Q_{jk}^{1-x_{jk}}(\theta)$$

5  $\theta$  固定、項目パラメータ所与とする。数式 1 2 の対数尤度を取り、類似項目  $j$  に係わる部分だけを取り出すと

【数 1 3】

$$\ell_{B(j)} = \sum_{k=1}^s (x_k \ln(P_k) + (1-x_k) \ln(1-P_k))$$

が得られる（ただし正答確率などに対する添え字  $j$ 、及び  $\theta$  は省略する）。すると、その正誤反応  $X$  に関する期待値は

10 【数 1 4】

$$E(\ell_{B(j)}) = \sum_{k=1}^s (P_k \ln(P_k) + (1-P_k) \ln(1-P_k))$$

となる。また

【数 1 5】

$$\bar{P} = \frac{1}{s} \sum_{k=1}^s P_k$$

15 【数 1 6】

$$P_k - \bar{P} = \Delta P_k$$

とするとき、 $\Delta P_k$  は、各項目が類似項目であるため微小と仮定する。 $\Delta P_k$  の 2 次以上の項を無視すると

【数 1 7】

$$\begin{aligned} P_k \ln(P_k) &= (\bar{P} + \Delta P_k) \ln(\bar{P} + \Delta P_k) \\ &\cong (\bar{P} + \Delta P_k) \left( \ln(\bar{P}) + \Delta P_k \times \frac{1}{\bar{P}} \right) \\ &\cong \bar{P} \ln(\bar{P}) + \Delta P_k \ln(\bar{P}) + \Delta P_k \end{aligned}$$

20

【数 1 8】

$$\begin{aligned}
 (1-P_k)\ln(1-P_k) &= (1-\bar{P}-\Delta P_k)\ln(1-\bar{P}-\Delta P_k) \\
 &\cong (1-\bar{P}-\Delta P_k)\left\{\ln(1-\bar{P})-\Delta P_k \times \frac{1}{1-\bar{P}}\right\} \\
 &\cong (1-\bar{P})\ln(1-\bar{P})-\Delta P_k-\Delta P_k \ln(1-\bar{P})
 \end{aligned}$$

が成立する。数式 17 及び 18 を数式 14 に代入し、

【数 19】

$$\sum_{k=1}^J \Delta P_k = 0$$

5 であることを考慮に入れば

【数 20】

$$\begin{aligned}
 E(\ell_{B(j)}) &\cong \sum_{k=1}^J \left\{ \bar{P} \ln(\bar{P}) + (1-\bar{P}) \ln(1-\bar{P}) \right\} + \ln\left(\frac{\bar{P}}{1-\bar{P}}\right) \sum_{k=1}^J \Delta P_k \\
 &= \sum_{k=1}^J \left\{ \bar{P} \ln(\bar{P}) + (1-\bar{P}) \ln(1-\bar{P}) \right\} \\
 &= s \left\{ \bar{P} \ln(\bar{P}) + (1-\bar{P}) \ln(1-\bar{P}) \right\}
 \end{aligned}$$

次に部分得点についても検討してみよう。先程と同様に全ての項目で繰り返しの回答  $s_j$  が等しい ( $=s$ ) と仮定した場合、数式 9 の正誤の確率を数式 15 の

10 平均反応確率と

【数 21】

$$\bar{Q}_j(\theta) = 1 - \bar{P}_j(\theta)$$

で置き換える。すると

【数 22】

$$15 \quad \ell_{part} = \sum_{j=1}^n s \left\{ r_j \ln(\bar{P}_j(\theta)) + (1-r_j) \ln(\bar{Q}_j(\theta)) \right\}$$

となる。上式より特定の類似問題  $j$  に係わる部分だけを取り出し、また部分得点  $r$  が類似問題  $j$  の正誤の平均であることから

【数 23】

$$\ell_{part(j)} = s \left\{ \frac{\sum_{k=1}^s x_k}{s} \ln(\bar{P}) + \left( 1 - \frac{\sum_{k=1}^s x_k}{s} \right) \ln(1 - \bar{P}) \right\}$$

となる（ただし正答確率などに対する添え字  $j$ 、及び  $\theta$  は省略する）。すると、その期待値は

【数 2 4】

$$\begin{aligned} E(\ell_{part(j)}) &= s \left\{ \frac{1}{s} \ln(\bar{P}) \times E\left(\sum_{k=1}^s x_k\right) + \ln(1 - \bar{P}) \times E\left(1 - \frac{\sum_{k=1}^s x_k}{s}\right) \right\} \\ &= s \left\{ \frac{1}{s} \ln(\bar{P}) \times \sum_{k=1}^s P_k + \ln(1 - \bar{P}) \times \left(1 - \frac{1}{s} \sum_{k=1}^s P_k\right) \right\} \\ &= s \{ \bar{P} \ln(\bar{P}) + (1 - \bar{P}) \ln(1 - \bar{P}) \} \end{aligned}$$

であり、数式 2 0 及び 2 4 から

【数 2 5】

$$E(\ell_{B(j)}) = E(\ell_{part(j)})$$

が成立する。尤度を構成する全ての類似項目に数式 2 5 が成立すれば、部分得点の最尤解と、2 値データによる解は近似的に一致すると考えられる。

なお、以上で、本発明における部分得点モデルは、同一パラメータの問題又は類似パラメータにおいて繰り返し定義の回数が同一であれば、通常の 2 値の項目反応モデルと対応関係を有することを既に示した。しかし、現実の適用場面では、すべての問題の繰り返し回数が同一である、つまり、部分得点の段階数がどの問題についても同一であるとは限らない。例えば、学力テストにおいては、初めに 2 値的に採点される易しい項目があり、後半に多値的に採点される難しい項目が配置されることが多い。このような状況に対応するためには、数式 9 を次のように拡張する必要がある。

【数 2 6】

$$\ell_{part} = \sum_{j=1}^n s_j (r_j \ln(P_j(\theta)) + (1-r_j) \ln(Q_j(\theta)))$$

このような拡張型部分得点モデルでも、一般の 2 値モデルの最尤解と数式 2 6 の解とは近似的に一致する。

次に、2 項分布を利用した繰り返し回数  $s_j$  の推定について説明する。拡張され

## 5 た部分得点モデル

### 【数 2 7】

$$\ell_{part}(\theta) = \sum_{j=1}^n s_j (r_j \ln(P_j(\theta)) + (1-r_j) \ln(Q_j(\theta)))$$

を適用する場合、各項目の繰り返し回数  $s_j$  はシミュレーションでもない限り実際には不明のことが多く推定する必要がある。部分得点の性質により  $s_j$  の推定は異

## 10 なり、以下で説明を加える

第 1 に、部分得点が複数の正誤問題の平均値である場合について考察する。ある問題 j の部分得点が、1 組の組問などからなる複数の正誤 2 値問題の平均値（つまり複数の問題の合計点を採点結果とする場合）である場合は、それらの問題の数が、そのまま繰り返し数  $s_j$  の推定値となる。

15 第 2 に、部分得点が質問紙などの段階評定である場合について考察する。質問紙で多用される評定尺度の場合、段階数を  $m+1$  とするとき繰り返し数  $s$  は  $m$  と推定できる。たとえば質問項目が、「1. よくあてはまる」「2. ややあてはまる」「3. ややあてはまらない」「4. まったくあてはまらない」の 4 段階の評定で回答を求めるものであるならば、回答結果を 0～1 の部分得点として処理す

るために、各回答を  $x$  とするとき  $(x-1)/3$  と変換して「0, 1/3, 2/3, 1」の 4 段階をとる部分得点データとして分析することになる。これは、3 つの 2 値項目の得点の和を 3 で割ったことと同じである。つまり質問紙の回答が  $m+1$  段階の評定であるならば、2 値データの繰返し数  $s_j$  は  $m$  であると推定できる。

- 5 第 3 に、部分得点が、複数の正誤 2 値問題の平均値でなく、採点者の評定のよ  
うな採点結果であり、比較的段階数が少ない場合について考察する。最も簡単な  
 $s_j$  の推定方法は、部分得点  $r$  が  $s+1$  段階の得点であるならば、元の 2 値データの  
項目の繰返しは  $s$  であると推定するものである。部分得点モデルの考えでは、  
2 値データの合計を繰返し回数で割ったものが部分得点  $r$  と対応していたこと  
10 から、たとえば、採点者の評価結果が 0、0.5、1 の 3 段階で表現されるので  
あれば、これは 2 つの 2 値項目の得点の和を 2 で割ったことに他ならない。もち  
ろんこれが成立するためには、得点の発生する可能性のある各段階間の等間隔性  
が必要となる。同じ 3 段階の部分得点であっても 0、0.5、1 と 0、0.75、  
1.0 の持つ意味は異なる。部分得点が同一母数項目の繰返しにより生ずるとい  
15 う仮定からは繰返し数 2 で後者のパターンが生じるとは考えにくく、より大きな  
繰返し数を仮定する必要があるだろう。たとえば 0、0.25、0.5、0.75、  
1.0 のパターンを想定し、繰返し 4 を考える場合である。

- 第 4 に、部分得点が複数の正誤 2 値問題の平均値でなく、採点者の評定のよう  
な採点結果であり、段階数が多い場合について考察する。既に述べたように部分  
20 得点  $r$  が  $m+1$  段階であるとき 2 値の正誤項目からこれを構成するためには繰返



し回数  $s$  は  $m$  であることが必要である。段階数が少ない場合は、前節の推定で良いと考えられるが、段階数が多くなると問題が発生する可能性が高くなる。たとえば、100点満点のテストで、1人の結果を除いた解答に採点者が10点刻みに採点した場合、段階数は11段階で繰返し数は10である。しかし採点者が残る1人の解答結果に69点という点をつけたとするとその途端に、繰返し数を101としないと部分得点化が困難となる。また得点を整数に限らず実数も許容する場合には、このような考え方では繰返し数を推定することは困難となる。

以上のような問題を解消するため、ここでは若干の仮定を設けた上で繰返し数を推定する。問題  $j$  の部分得点は、当該の問題に関係する潜在的な問題の繰返し実施による正誤の平均で部分得点が構成されることを仮定し、正答確率は  $P_j$ 、誤答確率は  $Q_j = 1 - P_j$  とする（以上部分得点の仮定）。このとき  $\theta$  を固定すれば、部分得点  $R$  が  $0, 1/m, 2/m, \dots, 1$  をとる確率  $P_m(R=r|\theta)$  は2項分布

【数28】

$$P_m(X=x|\theta) = {}_m C_x P_j^x Q_j^{m-x}, \quad x=0, 1, \dots, m$$

に一致する。すると母集団全体では、 $m$  回の繰返しで部分得点が  $r$  となる確率は、

【数29】

$$P_m(R=r) = \int {}_m C_x P_j^x Q_j^{m-x} \varphi(\theta) d\theta$$

であり、その分布関数は

【数30】

$$F(r) = P(R=0) + P(R=1/m) + \dots + P(R=r)$$

となる。ここで  $\varphi(\theta)$  は  $\theta$  の母集団分布である。この理論分布とデータから得られる経験分布が一致することが望ましい。このため理論分布と経験分布の一致度をコルモゴロフ・スミルノフ検定の統計量

【数31】

$$KS\text{値} = \max |F(r) - F'(r)|$$

で評価する。ここで  $F(r)$  は、理論分布における段階  $r$  までの分布関数の相対累積度数であり、 $F'(r)$  は経験分布における段階  $r$  までの相対累積度数である。

実際にコルモゴロフ・スミルノフ検定の統計量を求めるためには、理論分布、経験分布とも  $r$  が離散量であり、また両方で段階の値が必ずしも一致しないこと

から生じる比較の困難さがあり、ここでは次のような方法で比較することにした。

まず  $\varphi(\theta)$  を標準正規分布と仮定する。もちろんこれは必要に応じて  $\theta$  の母集団分布と想定される任意のものをを用いることが出来る。これにより、数値積分を行えば数式 29 が求まり、その結果数式 30 が分る。R は 0、 $1/m$ 、 $2/m$ 、 $\dots$ 、

5 1 の  $m+1$  段階の値を取り得るわけだが、R が潜在的には連続量であると仮定すれば、たとえば  $\ell$  番目の段階値  $\ell/m$  の度数は、区間  $\left[\frac{\ell}{m} - \frac{1}{2m}, \frac{\ell}{m} + \frac{1}{2m}\right]$  に含まれる

R の度数の反映に他ならない。このため、 $\ell$  番目の段階値  $\ell/m$  までの理論分布関

数  $F\left(\frac{\ell}{m}\right)$  と比較する経験分布関数  $F'\left(\frac{\ell}{m}\right)$  は、部分得点としての区間  $[0, 1]$  の内

の区間  $\left[0, \frac{\ell}{m} + \frac{1}{2m}\right]$  の相対累積度数とすることにする。

10 以上のような方法で、最小のコルモゴロフ・スミルノフ検定の統計量を与える段階数  $m$  を、繰返し数  $s_j$  の推定値とすることが可能であり、以下のシミュレーション及び実際のデータへの適用により、この方法の有効性が確認されている。

第 5 に、シミュレーションによる繰返し数  $s_j$  の推定法の検討について述べる。

問題  $j$  の部分得点  $r_j$  が正誤 2 値問題の繰返し提示による正誤平均として表現されることを仮定して、シミュレーションによりこれを再現し、繰返し数  $s_j$  として  
15  $s_j = 5, 10, 20, 40$  の 4 種類（1 つの問題の部分得点データを作るために 2 値の問題を 5 ～ 40 個利用したということ）を行い、40 問題 3000 人分のデータを作成した。

前節の繰返し数  $s_j$  の推定法は、項目母数を与えられていなければならない。こ

20 のため作成された部分得点データに基づいて項目母数と能力母数の同時推定を行った（この段階での繰返し回数  $s_j$  は 1）。

図 7 a は、真の  $s = 5$  のとき、推定された経験分布と理論分布の分布関数の差の最大値（コルモゴロフ・スミルノフ検定の統計量）を、繰返し数  $s = 3 \sim 10$  についてプロットしたものである。ここで問題にしているのは繰返し数が多い

25 ときのケースであるから、繰返し数  $s = 1, 2$  は、元々不要であるので省略している（2 項分布を利用した繰返し数の推定は最低でも  $s = 3$  からとしたい）。こ

の図より明らかに真の繰返し数  $s$  の点でコルモゴロフ・スミルノフ検定の統計量が最小値をつけていることが分る。図 7 b には  $s = 10$ 、図 7 c には  $s = 20$ 、図 7 d には  $s = 40$  の場合を示したが、いずれも真の繰返し数で最小値をとっていることが分る。

- 5 第 6 に、能力検査への適用例として、実際のデータでの繰返し数の推定例について述べる。英語力テストでの繰返し回数  $s_j$  の推定への適用例を示す。同テストは受験者 9480 人、問題数 43 項目のデータであり、長文読解に絡んで複数の小問をあてているため 5 問組が 2 組存在している。両組の小問の正誤平均で部分得点を作成し、前節の繰返し数の推定を試みた。その結果得られたのが図 8 a  
10 及び図 8 b である。どちらも当該群の問題数の 5 でコルモゴロフ・スミルノフ検定の統計量は最小値をとっていることが分り、推定の有効性が示されている。

以上では、本発明によるコンピュータ適応型テスト設計及び処理システムにおいて用いられる部分得点モデルについて説明した。この部分得点モデルは、2 値の項目反応モデルを修正することによって得られている。従って、この部分得点  
15 モデルでは、推定すべきパラメータの数も 2 値モデルの場合と同じであり、問題の特性に特別な関心があるとき以外は、学力テストの設計及び結果処理の際に、従来試みられてきたような複雑な多値モデルを利用する必要性は少ないといえよう。

本発明において用いられる部分得点モデルを従来から公知であった段階反応モデルなどと比較すると、次のようなことが明らかとなる。(1) 項目反応理論では、通常はデータが 2 値で 1 次元であることを要求しているが、部分得点モデルは、多値、多次元データに適用できる。(2) 部分得点モデルは(段階反応モデルや他の多値モデルに比較して)簡明なモデルで利用者に理解しやすい。(3) 部分得点モデルは、(段階反応モデルや他の多値モデルと違い)一般によく利用  
20 されている 2 母数ロジスティックモデルとシームレスなので、結果の解釈も容易であり、2 値と多値の混在データの分析にも便利である。(4) 部分得点モデルは、モデルのパラメータの数が(段階反応モデルや)他のモデルに比較して少なく、推定に問題が生じない。(5) 部分得点モデルは、どのような解答(回答)結果でも 0 から 1 の部分得点に変換すれば適用できるので応用範囲が広い。

(6) 部分得点モデルは、テストだけでなく質問紙データにも容易に適用できる。

(7) 部分得点モデルは、解答（回答）の段階数が増えても（段階反応モデルや他のモデルのように）モデルのパラメータが増加しないので、採点結果の変更に対応が容易である。

5      また、本発明の発明者の1人（藤森）は、部分得点モデルを用いたシミュレーションを行ったが、その結果によると、(1) 段階反応モデルは、テストが少数項目であると識別力の推定にバイアスがかかるが、部分得点モデルではこの現象は生じない。(2) 正答数得点と能力推定値の順位相関は、部分得点モデルの方が段階反応モデルより高い（真の値との相関では両モデルともほぼ同等）。

10      以上で本発明によるテスト・システムにおいて用いられる部分得点モデルの概要を説明したが、その要旨を再確認しておきたい。本発明が依拠する部分得点モデルにおいては、合計点を和を取った項目数で割ることにより0から1の間の部分得点として入力データとする、すなわち、2値モデルの尤度の正誤に使われるダミー変数を部分得点を表す実質的変数あるいは正誤確率の重みとして扱う。この  
15      ような想定の下で部分得点を処理しても、理論的にも従来の2値型の処理の場合との整合性が保たれ、また、シミュレーションの結果を見ても2値データの場合の尤度関数を用いる推定結果と部分得点モデルによる尤度関数を用いる推定結果とは実質的に同じであるといえる。

20      次に、上述した部分得点モデルが本発明によるテスト・システムにおいてどのように具体的に適用されるかを説明したい。具体例として、英語能力テストを考える。多肢選択式など解答が正答又は誤答の2値で評価される設問に関しては、従来型の項目反応理論が適応され、数式4の尤度関数 $L_j$ を用いて項目パラメータと能力値 $\theta$ とが推定される。これに対し、ライティング（作文）やスピーキング（会話）の問題では、正誤判断による2値的な採点ではなく部分点を許容する  
25      採点を行うことが、受験者の英語能力評価をより分析的かつ精密に行うことを可能にすると思われる。そのような場合には、2値評価を前提とする従来型の項目反応理論を適用することはできず、上述した部分得点モデルが用いられる。

部分得点モデルでは、例えば作文問題に対する受験者の解答を採点する際に、零点（0）から満点（1）までの間で25%刻みの部分得点を与えることができ

る。具体的には、数式 5 において  $r_{ij}$  が、0、0.25、0.50、0.75、1 という 5 つの値を取りうる尤度関数を考えて、部分得点を処理する。P を定義する数式 3 に含まれる項目パラメータ a (識別力) 及び b (困難度) は、同じ問題を用いて予め行われたプレテストにおけるデータを用いて既に推定されている。

- 5 既に一般論として論じたように、本発明の部分得点モデルでは、部分得点を同一パラメータを有する問題が反復的に出題された場合の正誤の平均として表現されるものと想定しているため、1 つの項目 (= 問題) の内部では、どの部分得点についても項目パラメータ a 及び b の値は同じである。

- 10 具体例として、このような英作文の問題が 3 題出題され、ある受験者によって与えられた解答がそれぞれ 0.25、0.75、0.50 と評価される場合を考える。この場合、0 と 1 とを含めると 5 段階の部分得点を用いて評価されているので、繰返し数は 4 であり、尤度関数  $L_{\theta, a, b}$  の対数尤度を考え、数式 9 の  $r_{ij}$  に部分得点 0.25、0.75、0.50 を代入することによって次のように決定される。

15 【数 3 2】

$$\begin{aligned} \ell_{part}(\theta) &= \sum_{j=1}^n s_j (r_j \ln(P_j(\theta)) + (1-r_j) \ln(Q_j(\theta))) \\ &= 4 [ \{ (0.25) \ln P_1(\theta) + (0.75) \ln Q_1(\theta) \} + \\ &\quad \{ (0.75) \ln P_2(\theta) + (0.25) \ln Q_2(\theta) \} + \\ &\quad \{ (0.50) \ln P_3(\theta) + (0.50) \ln Q_3(\theta) \} ] \end{aligned}$$

- 20 ここで、ロジスティック曲線 P に含まれる項目パラメータ a 及び b は、それぞれの問題に対して予め推定されている。数式 3 2 の対数尤度を用い、最尤推定法やベイズ推定法により、受験者の能力  $\theta$  が推定される。これらの方法自体は従来から知られている一般的な統計学的手法であり本発明の特徴ではない。しかし、いずれの推定方法を用いるにしても必要な対数尤度は、本発明の核心である部分得点モデルによって初めて得られるものである。

25 以上は、部分得点モデルの基本的な適用例である。更に、部分得点モデルを用いると、このような単純な例とは異なり、複数の評価基準を設け、1 問を複数の視点から評価する分析的な評価方法を採用することによって、受験者の能力推定

の精度を向上させることもできる。例えば、上述した基本的な適用例では顕在化しない受験者の能力差を明らかにできる可能性がある。例えば、ライティング問題では、精度の高い評価のために、（a）目的達成の有無（Goal Achievement）、（b）文法（Grammar）、（c）語彙（Vocabulary）、

5 （d）文章構成（Style）、（e）綴り字・句読点（Mechanics）という5つの評価基準を設定し、スピーキング問題であれば、（a）から（c）に加え、（f）発話の流暢さ（Fluency）、（g）発音（Pronunciation）という5つの評価項目を設定する。このような評価項目を設定する理由は、例えば、作文の設問であれば、非常に拙い書き手の意図が一応は読み手に伝わる、文法的には

10 誤りはないが余りに不自然であり状況にそぐわない、明らかな文法上の誤りがある、単語の綴りに間違いがある、類似する単語が誤用されている、文の途中で中断し完結していない、パンクチュエーションが間違っているなど、ただ1つの作文を評価する場合でも、異なる複数の視点から評価を行えば、1つの問題を複数の小問題から構成されているものと考え、視点ごとに異なる評価が可能である

15 からである。そして、このように複数の評価項目を設定すると、1つの問題を複数の問題から構成されているものと見なすことによって、上述した基本的な適用例の場合よりも高精度の評価が可能となりうる。

以上のような評価項目を設けた場合には、次のような評価が可能であろう。

（1）の目的達成の有無については、例えば、「相手に宿を手配して欲しいという意図の伝達」という目的が問題ごとに設定され、設定された目的が「達成されている、達成されていない」の2値（1又は0）で採点される。それ以外の評価項目に関しては、0から1までの間に25%刻みの部分得点を許容する採点を行う。つまり、数式5の $r_{ij}$ に、0、0.25、0.5、0.75、1という5つの値が代入される。上述した基本的な適用例と同様に、尤度関数に含まれるロジスティック曲線Pを定義する数式3に含まれる項目パラメータa（識別力）及びb（困難度）も、上述の例と同様に、同じ問題を用い本番のテストに先行して行われたプレテストにおけるデータを用いて予め推定されている。このような採点方法を用いれば、ただ1題の英作文又は会話の問題を出題することによって、非常に多角的な評価が可能になる。ただし、ここで挙げた評価項目は単なる例示

20

25

であり、本発明によるテスト・システムにおいて他の視点から評価を行うことも可能である。

以上が、部分得点モデルを含む項目反応理論が英語能力テストへ適用され受験者の能力が推定される様子についての説明であり、本発明は、このような部分得点モデルを含む項目反応理論による能力推定をインターネット接続環境にある一般的なパーソナル・コンピュータを用いて実現するテスト・システム及びテスト方法である。以下では、添付の図面を参照しながら、本発明によるテスト・システムの動作を概説する。

図 1 には、本発明によるテスト・システムの第 1 の実施例の概要が示されている。受験者は、本発明によるシステムによって設計、実施及び処理されるテスト（例えば、英語能力テスト）を実施する語学学校などに設置されたインターネット接続環境にあるパーソナル・コンピュータである受験者ユニット 101 を用いてテストを受験する。受験者の認証が適切になされるのであれば、自宅での受験も可能である。受験者によってキーボード、マウス、マイクロフォンなどを介してパーソナル・コンピュータである受験者ユニット 101 に入力される解答結果は、インターネットなどのネットワーク 103 を介して、採点者ユニット 102 に送られ、例えばテスト対象である英語を母語とする採点者が部分得点を許容しながら採点を行う。その際に、部分得点を与える評価項目は、問題の難易度にあわせて変更されることはない。ただし、解答結果は、受験者ユニット 101 から直接に採点者ユニット 102 に送信されるのではなく、このテストを集中的に管理するテスト管理サーバ 104 にいったん送られた後で、複数存在する中から適切であると判断される採点者ユニット 102 へ送られるのが一般的である。テスト管理サーバ 104 は、問題データベース 105 を備えている。問題データベース 105 には、プレテストとして実施され項目パラメータ（数式 3 における識別力  $a$  及び困難度  $b$ ）が予め推定されている問題群が記憶されている。テスト管理サーバ 104 は、受験者が特定されると、問題データベース 105 から一群の問題を選択し、選択された問題群を受験者ユニット 101 へ送信する。

図 2 には、受験者ユニット 101 の概要が図解されている。受験者ユニット 101 は、通常、インターネット接続環境を有する一般的なパーソナル・コンピュ

ータである。入力装置 207 は、キーボード、マウス、タッチパネルなど機械的な入力装置であり、音声入出力装置 209 は、マイクロフォンやスピーカなどである。受験者は、まず、入力装置 207 から手動で、又は、音声入出力装置 209 から音声で、自分自身の ID を入力すると共に、テストの開始を指示する。受験者の ID は、その受験者が受験の申込みを行う登録時に、テスト管理サーバ 104 から各受験者に対して一意的に発行されたものが用いられる。セキュリティ管理のため、ID と共にパスワードも発行されるのが一般的である。一意的な ID が発行されていることにより、同一の受験者が複数回受験する場合には、その旨がテスト管理サーバ 104 において認識され、適切な出題がなされる。指示に  
5 応答してテスト管理サーバ 104 から送信されディスプレイ 208 に表示される、又は、スピーカを含む音声入出力装置 209 から出力される自分のレベルに合わせて選択された問題に対して、受験者は、入力装置 207、又は、マイクロフォン（音声入出力装置 209）を介して解答を入力する。解答、特に、部分得点を許容する採点が必要となるライティングやスピーキングの設問に対する解答は、  
10 通信インターフェース 202 とインターネットなどのネットワーク 103 とを介して、採点者ユニット 102 に送られる。ただし、既に述べたように、解答は、受験者ユニット 101 から直接に採点者ユニット 102 に送信されリアルタイムに採点されるのではなく、このテストを集中的に管理するテスト管理サーバ 104 にいったん送られた後で、複数存在する中から適切であると判断される採点者  
15 ユニット 102 へ送られるのが一般的である。一定数の解答が集まってから採点するのが効率的であるという採点の経済からもこれは当然であろう。

図 3 には、採点者ユニット 102 の概要が図解されている。採点者ユニット 102 は、受験者ユニット 101 と同じように、通常、インターネット接続環境を有する一般的なパーソナル・コンピュータである。インターネットなどのネット  
20 ワーク 103 を介して受験者ユニット 101 又はテスト管理サーバ 104 から送られてくる解答結果をディスプレイ 308 に表示し、又は、スピーカ（音声入出力装置 309）から出力して、キーボードやマウスなどの入力装置 407 を用いて採点する。採点結果は、インターネットなどのネットワーク 103 を介してテスト管理サーバ 104 に返送する。



以上で説明した本発明のテスト・システムを実現する第１の実施例では、受験者ユニットと採点者ユニットとテスト管理サーバとがインターネットなどの通信回線を用いた通信ネットワークの通信端末として構成されていた。しかし、当業者には明らかなことであるが、本発明によるテスト・システムは、第２の実施例として、通信機能を備えていないスタンドアロンのパーソナル・コンピュータを用いて実現することもできる。その場合、難易度及び識別力が予め推定された多数の問題が記憶されているデータベースは、例えば当該パーソナル・コンピュータのハードディスクなどの記憶装置に構築されており、受験者は、例えばＣＤやＤＶＤなどに格納された態様で提供される本発明のテストを実施するプログラムに含まれる命令に従って出題されるライティングやスピーキングの問題に対し、キーボードやマイクロフォンを介して解答する。解答結果はいったんハードディスクなどに記憶され、採点者は、その解答結果をハードディスクから読み出して部分得点を許容する採点を行う。この第２の実施例の場合の部分得点の処理方法は、第１の実施例の場合と同様である。採点がなされると、部分得点モデルに基づく尤度関数を用いて受験者の能力推定が行われる。

図４は、本発明によるテスト・システムを用いたテスト実施の概要を示す流れ図であり、特に、第１の実施例に則して、本発明が依拠する部分得点モデルが用いられるライティング及びスピーキングに関するテストの実施及び処理プロセスが示されている。なお、ライティングの問題では、受験者は、キーボード（図２の入力装置２０７）を用いて、英文などの文章をタイプインする形式で解答するのが一般的である。スピーキングの問題では、受験者がパーソナル・コンピュータに備え付けられたマイクロフォン（図３の音声入出力装置３０９）を介して、提示された質問に対する応答として、又は、自由な内容で音声を入力し、その音声として発話された内容が評価の対象となる。採点者は、受験者が現在利用しているパーソナル・コンピュータとインターネットなどのネットワークを介して接続されている別の端末である採点者ユニットの前で待機し、リアルタイムで採点を行うことも可能ではあるが、実際には、受験者の解答は、いったんテスト管理サーバに記憶された後で採点者ユニットへ送られ、集合的に多数の解答を採点するのが一般的である。

まず、受験者は、受験者ユニット１０１においてインターネット上の指定されたウェブページにアクセスする。そのウェブページにおいて、自分自身の受験者ＩＤとパスワードとを入力し認証がなされた後で、所定のボタンをクリックすることにより、ライティング又はスピーキング問題の受験を開始する意図を通知する（４００）。テスト管理サーバ１０４は、この通知に応答して、ライティング又はスピーキングの問題を問題データベース１０５から選択する（ステップ４０１）。この選択の際には、この受験者による多肢選択式問題に対する解答の採点結果から推定された能力 $\theta$ との関係で、最も適切な識別力及び困難度を有する評価項目を含む問題を選択することができる。例えば、ライティング問題選択の際にはリーディング問題への解答結果を参考にし、スピーキング問題選択の際にはリスニング問題への解答結果を参考にすることが考えられる。これは、ライティングの能力はリーディングの能力と相関関係を有し、スピーキングの能力はリスニングの能力と相関関係を有すると考えられるからである。ただし、このような問題選択は、単なる例示であり、本発明による部分得点モデルを含む項目反応理論に基づくテスト・システムの本質部分ではない。

選択された問題に対応するロジスティック曲線に含まれる項目パラメータは、先に実施されたプレテストにおけるデータから、評価項目ごとに予め決定されているのが一般的である。ただし、能力パラメータの推定と同時に項目パラメータを推定する同時最尤推定なども理論的には可能であり、本発明によるテスト・システムも、この同時最尤推定の可能性は排除しない。その場合には、問題データベース１０５に記憶されている問題の項目パラメータは既に推定されている場合とそうでない場合とが併存することになる。項目パラメータの推定は、能力 $\theta$ の推定と同じく数式５の尤度関数が用いた部分得点モデルに基づいて行われる。また、項目パラメータが推定される際には、それぞれの問題の識別力及び困難度を標準化する等化と称されるプロセスも行われる。この等化のプロセスにより、受験者の母集団に左右されない絶対評価が可能になる。ただし、等化自体は項目反応理論一般について妥当することであり、特に本発明の特徴ではない。

テスト管理サーバ１０４によって出題されるべき問題が選択されると、選択された問題は、インターネット等のネットワーク１０３を介して受験者ユニット１

01に送信される(ステップ402)。送信された問題は、テキスト形式の問題であればディスプレイ208において、音声形式の問題であればスピーカ(音声入出力装置209)から受験者に対して出題される(ステップ403)。受験者は、その出題された問題に対して、文章をタイプインする形式で、又は、発話された音声形式で、解答を与える(ステップ404)。解答を構成する文書又は音声ファイルは、インターネット等のネットワーク103を介してテスト管理サーバ104に送信され、いったん記憶される(ステップ405)。

以上のプロセスが一定数の受験者に対して反復され、一定数の解答ファイルがテスト管理サーバ104に記憶される。ただし、これら複数の受験者に対して出題される問題は、同じとは限らない。項目反応理論の一般論から、出題される問題は異なっているとしても、能力 $\theta$ は適切に推定されうるからである。採点者ユニット102からテスト管理サーバ104により開設されているウェブページへのアクセスがなされ、テスト管理サーバ104に蓄積されている解答を採点のために送信して欲しい旨のリクエストがなされると、一定数の解答ファイルが、採点者ユニット102に送られる(ステップ406)。採点者は、既に説明した部分得点を許容する方法により解答を採点し(ステップ407)、採点結果をテスト管理サーバ104に返送する(ステップ408)。ただし、一般的には、このように複数の解答がまとめて採点されるか、リアルタイムで採点されるかは、本発明の特徴とは関係ない。

次に、テスト管理サーバ104は、採点者ユニット102から受信した部分得点を数式5の尤度関数に代入し、能力 $\theta$ の推定を行う(ステップ409)。上述したように、項目パラメータが同時に推定される場合もありうる。推定の方法は、最尤推定、ベイズ推定などである。推定が終了すると、必要な場合には、推定された $\theta$ の値が、他のテストとの比較に適した点数に換算される(ステップ410)。

本発明のテスト・システムを用いて上述のような処理を行うことにより、従来から行われてきた一般的な項目反応理論では処理が困難であった部分得点を許容して採点を行う問題を含むテストであっても、従来型の項目反応理論と同様な能力推定を行うことができる。

発明者らは、本発明によるテスト・システムを用いた能力推定の有効性を確認するために、2003年5月21日から28日において、被験者12名によるスコア安定性確認調査を行った。方法としては、同一被験者に3回続けて本発明によるテスト・システムを用いた英語能力テストを受験してもらい、スコアに大きなぶれが生じないかどうかを確認した。ここで実施された英語能力テストは、リスニング、リーディング、ライティング、スピーキングの4つの技能のテストで構成されていた。

被験者は、東京の比較的英語が得意な某大学の大学生12名である。もし、本発明によるテスト・システムによる英語能力評価が適切なものであり、等化を含めた各問題の項目パラメータの推定が本発明によるテスト・システムにおいて適切になされているのであれば、同一受験者が1日のうちに3回テストを受験した場合、英語能力はその間に変化しないと考えられるので、結果のスコアは大きく変動しないはずである。

結果の詳細は、図5のグラフとして示されている。また、12名の平均スコアは、図6の通りである。個人差はそれぞれあるが、図5及び図6において観察されるように、1回目から2回目において試験形式の慣れによる得点の向上（一般的には「リハーサル効果」と呼ばれる）と、3回目の受験における疲労効果（特に、最後の回の最終問題であるスピーキングの得点に見られる）による得点の低下とが見られた。

しかし、平均点で見ると各技能で250点満点中5.4から18.5点のスコアの変動（＝2.2％から7.4％）であり、トータルスコアでも3％の変動幅で収まっている。従って、異なった問題を受験しても本発明によるテスト・システムでは、テスト・スコアの変動は少なく、スコアの標準化が適切に行われており、母集団の能力レベルに左右されない絶対評価が可能であることがわかった。

同一パラメータを有する複数の項目が反復的に出題された場合の正誤の平均として部分得点が表現されると想定すれば、理論的には、数式5における $r_{ij}$ を実質的な変数と考え0から1までの間の部分得点を許容しても、従来型の項目反応理論における2値的な評価に用いられる尤度関数の場合と推定結果は同一になる、というのが、本発明によるテスト・システムが依拠する部分得点モデルによ

る提唱であった。図 5 及び図 6 において示された実験結果は、この理論的帰結が、実験的にも確認されたことを示している。

5 従って、本発明によるテスト・システムによれば、従来型の項目反応理論との整合性を維持しながら、従来よりも高精度の能力推定が達成される。これは、本発明の著しい効果である。

10 明細書の最後に、本発明の発明者の 1 人（藤森）による段階反応モデルと部分得点モデルとの比較に関する論文（『項目反応理論による多値データの分析について－段階反応モデルと部分得点モデル－』）を、文中で言及されている図面（Figure）や表（Table）を除いて、以下に引用する。なお、この論文は、この出願における優先権主張の基礎となる日本特許出願である特願 2 0 0 2 - 2 7 9 5 1 3 が出願された平成 9 月 2 5 日の時点では未発表であったが、その後、文教大学人間科学部の紀要である『人間科学研究』2 4 号において発表された（平成 1 4 年）。また、この論文は、学力テストだけを念頭においたものではなく、より広い応用についても言及しているので、「問題」ではなく「項目」、「受験者」ではなく「被験者」など、以上の用例と異なる箇所が含まれている。

#### 1. 研究の目的

20 2 母数ロジスティックモデル（Birnbaum, 1 9 6 8）のような一般の項目反応理論では、正答、誤答のような 2 値データを処理することが出来るだけであり、多値のデータを分析することは出来ない。しかし心理学一般の研究では質問紙データなどで多値データを利用することも多い。項目反応理論でも多値データが全く扱えないわけではなく、幾つかの分析モデルが開発されている。例えば Samejima（1 9 6 9）の段階反応モデル（graded response model）、評  
25 定尺度モデル（rating scale model; Andrich, 1 9 7 8）や partial credit model（Masters, 1 9 8 2）はこの種の代表的なものとされる。しかし現実問題としては分析するためのソフトウェア環境が十分整備されていないこともあってか、これらのモデルが広く一般に利用されているとは言い難い。

このような状況に鑑み、本研究では藤森（2 0 0 1）の部分得点モデル（partial test score model）と段階反応モデルとの比較をシミュレーション及び質問紙データの分析結果をもとに行い、項目反応理論を利用した多値デ

ータに関する分析方法の検討を行うことにする。

多値データへ項目反応理論を適用するとしても、どのようなモデルが良いのか検討する必要がある。このためには（１）理論的側面の検討、モデル母数の再現性や能力母数の分布などの影響を検討するための（２）シミュレーションによる  
5 検討、そして現実のデータにどの程度うまくあてはまるのかという（３）実証データに基づく検討が求められるであろう。このため、本研究では（１）を１．３節で、（２）を２節で、そして（３）を３節で行うことにする。

### １．１．段階反応モデル

m個の順序性を持つ段階反応を許容する Samejima (1969) の段階反応モデルを初めに説明する。ここではテストのある一つの項目だけを考えることにする。被験者の回答の各段階に対応して決まるダミー変数を  $u$  とし、ある段階より低い段階に反応したとき 0、そうでないとき 1 となるダミー変数を  $x$  とする。  
能力  $\theta$  を持つ被験者が  $x=1$  となる確率を

【数 3 3】

$$15 \quad P_u^*(\theta) = P(X=1|\theta)$$

で表すとき

【数 3 4】

$$P_u(\theta) = P_u^*(\theta) - P_{u+1}^*(\theta)$$

【数 3 5】

$$20 \quad P_u(\theta) > 0$$

が段階反応モデルである。ただし

【数 3 6】

$$P_0^*(\theta) = 1$$

【数 3 7】

$$25 \quad P_m^*(\theta) = 0$$

である。数式 3 3 は段階の間を確定する境界反応曲線であり、数式 3 4 は段階の

反応確率を表現する段階反応曲線となる。前記の条件を満たす限り数式 3 3 の関数はどのようなものでも自由であるが、本研究では良く利用される 2 母数ロジスティックモデル

【数 3 8】

$$P_u^*(\theta) = \frac{1}{1 + \exp(-a(\theta - b_u))} \quad u = 1, \dots, m-1$$

を仮定する。数式 3 5 式の条件のため、 $a$  は数式 3 8 の全ての曲線に共通した母数となり識別力と呼ばれる。 $b_u$  は各段階の閾値に関係した母数であり困難度と呼ばれる。すなわち段階反応モデルでは項目ごとに 1 つの識別力と各段階の閾値に対応する  $m-1$  個の困難度母数を持っていることになる。

#### 10 1. 2. 部分得点モデル

部分得点モデルでは部分得点を  $x$  ( $0 \sim 1$  の範囲) とするとき母数推定のための対数尤度は次式で表される。

【数 3 9】

$$\ell = \sum_{i=1}^N \sum_{j=1}^n s_{ij} \{ r_{ij} \ln(P_j(\theta)) + (1 - r_{ij}) \ln(Q_{ij}(\theta)) \}$$

15 ここで  $i$  は被験者、 $\theta$  はその特性値を表す母数、 $j$  は項目、 $s$  は 2 値項目換算の繰返し数、 $P$  は 2 母数ロジスティックモデル、 $Q = 1 - P$  である。

すなわち部分得点モデルは 2 母数ロジスティックモデルをその基礎に置いている。また同一あるいは類似項目母数を持つ項目を被験者に対し潜在的に繰返し実施することを想定している。このとき  $r$  は繰返し実施の正誤平均と考えることができる。このような部分得点モデルと繰返し実施を考えた 2 母数ロジスティック  
20 モデルの最尤解は一致する（類似母数の時は近似）ことが証明できる（藤森, 2002 a）。

#### 1. 3. 理論的側面からの検討

部分得点モデルの特徴は、以下のようなものである（藤森, 2002 b）。

- 25 (1) 項目母数は識別力と困難度で一般の利用者にも比較的理解しやすい。
- (2) モデル母数の数が他の多値モデルと比較して少なく推定上の問題がない。
- (3) どのような解答／回答結果でも値を  $0 \sim 1$  の部分得点に変換すれば適用で

きるので応用範囲が広い。(4) 質問紙の評定尺度にも容易に適用できる。

(5) 解答の段階数が増えても母数が増加しないので、採点結果の変更や微調整にも対応が容易である。(6) 2母数ロジスティックモデルとシームレスなので識別力や困難度母数の解釈もそのまま2母数ロジスティックモデルと同様に行えるので、2値と多値の混在データなどでの運用に便利である。(7) testlet/組間などの部分得点化により2値項目からなるテストに組間が存在する場合、組間を1つの部分得点項目としてとしての運用することが可能になり、副次的にテスト全体での母数の数が減る。(8) 残差得点にモデルを再適用して多次元データへの対応もできる。

10 (7) (8) については本研究では取り上げていないが特徴としては大きなものである。

続いて段階反応モデルの特徴は、以下のようなものである。(1) 項目反応理論で多値データに対応したモデルとして著名なものである。(2) 発表されてから30年以上の時間がたち応用研究も報告されている(例えば野口(1999)など)。(3) 公開されている分析用ソフトウェア MULTILOG が存在する。

15 (4) 各段階反応曲線が得られるため、各段階に対する反応にまで関心を持つ場合は便利である。例えば境界反応曲線の母数を利用すれば、段階ごとの反応が $\theta$ 上のどの水準で発生するのかについて予測も出来る。

両モデルを比較してみると、モデル化の考えの違いが明らかである。段階反応20 モデルは、モデルの母数の数が多いだけ、部分得点モデルに比較してデータとの当てはまりも良くなると考えられる。もちろんモデル母数の数が多ければ、母数の正確な推定にはデータ量が必要であるなどの問題が生じる危険性がある。逆に部分得点モデルは単純であるだけ、当てはまりは劣ると予想されるが推定値などの安定性は良いと考えられる。この点に関しては赤池情報量規準AICなどの観25 点が評価に必要となろう。

また段階反応モデルの(4)の点については、部分得点モデルでは、モデル上これは困難である。行うとすれば、事後的には段階ごとに被験者の分布を作成して検討することになるだろう。項目母数は所与であってもテスト実施の事前であれば、被験者の能力分布について適当なる仮定を設けて、シミュレーションを行



い段階ごとに被験者の分布を作成して検討することになるだろう。いずれにしても段階反応モデルより手間がかかることになる。

このような利点がある一方で段階反応モデルは（そして今まで提案された殆どの多値モデルも）、各段階にモデル母数を置くため段階の変化に柔軟性を欠く問題がある。たとえば、能力テストで教師が採点時に20点満点である項目を、0、5、10、15、20の粗い5段階で評価していたとしたら5段階の段階反応モデルでの分析となる。しかし、もし誤字脱字などである答案を1点減点することにしたなら、たちまちモデルの母数の値だけでなく、母数の数そのものも変化してしまうという問題点がある。質問紙でも評定を5段階で行っている項目を4段階にしたなら、段階反応モデルでは、それまでの項目母数をそのまま利用するわけにはいかなくなってしまふのである。

以上のように、どちらのモデルも一長一短はあるのだが、各段階や評定についてはそれほどの関心を持たず、項目の全体的な特性と能力母数の推定だけに関心がある場合には、部分得点モデルで十分と思われる。

## 15 2. シミュレーションによる検討

実際のテストや質問紙において回答がどのようなメカニズムで生起しているのかは正確に知ることは出来ない。部分得点モデルと段階反応モデルのどちらがより現実に近いかは現時点で判断することは難しい。このため両モデルの比較をシミュレーションで行うにしても、データの作成にあたっては、どちらかのモデルに従ってデータを作成すべきかを定めることは困難である。あるモデルを仮定してデータを作成し、当該モデルと別のモデルで分析しても、別のモデルは良い成績を発揮することは期待できず、比較もうまく行えないだろう。このため本研究では、部分得点モデルを前提としたシミュレーションと段階反応モデルを前提としたシミュレーションの2つを行うことにした。各モデルは自己のモデルに従ったデータと他のモデルのデータの2つを分析することになる。もちろん実際のデータはこの2者以外のメカニズムによる可能性もあるわけであるが、両モデルの公平な比較を行うことはできる。すなわちデータがモデルに従っている場合は、モデルの再現性がシミュレーションによって検討でき、またデータがモデルに従っていない場合はモデルの頑健性が評価できる。

## 2. 1. 部分得点モデルを前提としたシミュレーションデータ

部分得点モデルを前提としたシミュレーションデータは、以下のようにして作成した。まず部分得点モデルの構成要素として2母数ロジスティックモデルを仮定する。この2母数ロジスティックモデルの母数の分布型を以下のように定める。

5 識別力母数は、平均0.65、標準偏差0.25、下限0.3、上限2.0の切断正規分布、また困難度母数は、平均0、標準偏差0.5の正規分布に従うと仮定する。能力母数 $\theta$ は平均0、標準偏差1.0の正規分布に従うと仮定する。能力母数 $\theta$ を標準正規分布に従って作成し、2母数ロジスティックモデルから予想される正答確率を、範囲0～1の一樣乱数と比較し、前者が下回る場合被験者の

10 反応を正答1、上回る場合誤答0とする。2母数ロジスティックモデルに従う、この2値データパターンを、被験者数500、項目数200として各10回繰り返し作成した（データ1～10）。ただし5項目ずつ同一母数としている。続いて、このデータの同一母数の5項目ずつの正誤の和の平均をとり0、0.2、0.4、0.6、0.8、1.0の5段階の値をとる部分得点データとした。すなわち

15 部分得点データは、被験者数は2値データと同様に500であるが、項目数は40となる。同様にして、前記の項目母数を所与として新たに被験者の能力母数 $\theta$ を500人分追加してクロスバリデーション用のデータを作成した。

## 2. 2. 段階反応モデルを前提としたシミュレーションデータ

段階反応モデルを前提としたデータは以下のようにして作成した。段階反応モデルも、その構成要素として2母数ロジスティックモデルを仮定する。またその母数の分布形も0節と同様である。またデータの段階数としては1から5の値をとる5段階を仮定する。このため、モデルより段階間の境界反応曲線は4つ必要となる。この曲線群を確定するため、まず識別力母数を1つ分布に従って発生させ、これを各境界反応曲線に共通する識別力とする。続いて4つの困難度母数を作成し、最も小さいものを選択して、段階1と2の境界反応曲線の困難度とする。

25 以下同様にして困難度の小さなものから順にとり各境界反応曲線の困難度を決定する。これらの境界反応曲線間の差をとって各段階反応曲線とする。標準正規分布に従う能力母数 $\theta$ を1つ作成し、この値を固定して各段階反応曲線で予想される反応確率の区間（ $\theta$ を固定したときの全段階反応曲線の大きさの和は1であ

る)に、0～1の一様乱数が入ったとき当該反応が生じたことにする。以上の過程を500人分繰返し母数推定用のデータとした。また0節と同様に、以上で確定した項目母数を利用してクロスバリデーション用のデータを500人分作成した。

## 5 2. 3. 母数の推定

母数の推定は、両モデルとも項目母数と潜在特性値 $\theta$ の交互同時最尤推定による自作の FORTRAN プログラムによった。部分得点モデル用では交互同時最尤推定だけでなく項目母数の周辺最尤推定が可能であるが、段階反応モデルの推定プログラムは交互同時最尤推定のみに対応しているため、比較の便を考慮して両モデルとも項目母数を最尤推定した(結果は省くが部分得点モデルに関しては周辺最尤推定と交互同時推定の結果に大きな差はない)。両モデルとも $\theta$ と困難度の推定値は-3.5～3.5の範囲と定め、識別力の推定値の範囲は0.02～2.0としている。

## 2. 4. シミュレーションの結果と考察

15 表1は、部分得点モデルにより作成したシミュレーションデータ(以下部分得点データと呼ぶ)の能力母数の真値、正答数得点と両モデルで推定した $\theta$ の推定値の相関である。ただし相関は、ケンドールの順位相関係数を求めている(以下特に明示しない場合は相関はケンドールの順位相関を指す)。一般によく利用されるピアソンの積率相関でなくて順位相関を求めたのは、項目反応理論の現実の  
20 運用場面では正答数得点などと推定値の順位の逆転現象が問題となるケースが多いためである。表1より明らかなように真の $\theta$ との相関は、いずれも部分得点モデルによる推定値の方が高いものの段階反応モデルとの差はほとんどない。これに比し、正答数との相関は、若干差が拡大し部分得点モデルによる推定値が高い相関を与えている。

25 なお部分得点モデルの項目母数の推定値に関しては、例えばデータセット1については識別力の平均自乗誤差(MSE)は、0.0014、困難度に関しては0.0017であり、特に問題はない精度で推定値が得られている。ちなみに、データ作成の元となったモデルとは異なるので値の良し悪しは判断できないが、段階反応モデルの識別力との MSE は0.027、困難度は0.31となっている。

ただし困難度に関しては、段階反応曲線の困難度の推定値平均によって MSE を算出している。

続いて部分得点データの推定値を利用してクロスバリデーションデータの成績を比較してみよう（表 2）。クロスバリデーションにおいても様相は先と同様で、  
5 真値との相関はいずれも部分得点データがほんの僅か上回り、正答数得点に関してはややモデル間の差が拡大することが示されている。

さて段階反応モデルにより作成したシミュレーションデータ（以下段階反応データと呼ぶ）についての結果が表 3 である。また同データに基づき推定された項目母数をクロスバリデーションデータに適用した結果が表 4 である。どちらにおいても部分得点モデルに比べて段階反応モデルの方が真の  $\theta$  との相関がほんのわずかに高いが、大きな差ではなく、データセット 1, 5 のように逆転しているケースも見受けられる。また全体的に真値との相関が表 1 及び表 2 に比較してやや低くなっており、段階反応モデルが部分得点モデルに比較して再現性が難しいモデルであることを示唆していると言えよう。段階反応データであるにもかかわらず、  
15 正答数得点との相関は段階反応モデルより部分得点モデルによる推定値の方が高くなっている。しかも部分得点データの場合よりその差が拡大していることが分る。合計点との順序関係の逆転は段階反応モデルに多くなっているのであるが、この逆転が生じる理由は、段階反応モデルでは項目によって評定の間部分の段階反応曲線が相対的にかなり低くなり、他項目の回答結果によっては、当該項目  
20 の回答が  $\theta$  の推定に（推定誤差の大きさはともかくとして、その点推定値の決定には）殆ど影響力を持てなくなるためである。この現象は段階反応モデルだけではなく、選択肢やカテゴリーごとの反応確率曲線をモデル化する他の多値モデルにも共通するため、これらのモデルの利用の際には注意を払う必要があると考えられる。もちろん部分得点モデルも 2 母数ロジスティックモデルをベースとして  
25 いる以上、合計点と  $\theta$  の逆転も一部生じざるを得ないが、その程度は低く抑えられている。

次に段階反応モデルの項目母数の推定値について検討しよう。例えばデータセット 1 については識別力の平均自乗誤差（MSE）は、0.2993、困難度に関しては 0.0636 であり、やや部分得点モデルの場合に比較して推定精度が悪

いことがわかる。表 5 に示した項目 8, 15 はデータセット 1 の中の推定成績が悪い項目であるが、いずれも真の境界反応曲線が互いに接近しすぎている場合に推定が出来なくなっていることがわかる。このような場合には、2 つの境界反応曲線を 1 つとして処理するなどの対処が必要になると考えられるが、どのような基準で行うかなどの問題もあり本研究ではこれを行っていない。ある意味ではこのような対処を必要とすることに段階反応モデルの問題点が存在するとも言えよう。なおモデルは異なるものの部分得点モデルの識別力の MSE は 0.0136 となり段階反応モデルよりも良い推定値を与えている。ただし段階反応モデルの項目母数の推定に関しては周辺最尤推定では成績が改善する可能性も残されているので、ここではこれ以上取り上げず、別の研究で報告することにしたい。

以上をまとめると本研究のシミュレーションの条件下では、部分得点データ、段階反応データのいずれであっても真値との相関は、どちらのモデルを利用しても大きな差は生じていない。これに対し正答数得点との相関は明らかに部分得点モデルの方が良い成績であり、特に予期に反し段階反応データで差が大きくなっている。また項目母数の数が多い段階反応モデルで危惧されたクロスバリデーションデータでの不適合は生じていない（ただし母数推定上の問題は残る）。これは一つには、本研究が特にノイズをのせていない単純なシミュレーションであることも影響しているかもしれない。

### 3. 質問紙データによる検討

#### 3. 1. 質問紙データ及び母数の推定

本研究で分析の対象としたデータは、O 県の青少年基本調査の中高生 1849 人のデータで、教師と生徒の心理的距離について調べた 9 項目尺度である（表 6）。各項目は「1. よくあてはまる」から「4. まったくあてはまらない」の 4 段階で評定を求めるものであるが、9 項目とも回答は値が小さいほど心理的距離が小さくなるように処理されている。部分得点モデルでは、このデータを 0 ~ 1 の部分得点として処理するために、各回答結果を  $x$  とするとき  $x = (x-1) / 3$  と変換して「0, 1/3, 2/3, 1」の 4 段階をとる部分得点データとした。また推定方法はシミュレーションと同様の方法である。なお本調査の詳細については木原ら（1997）参照されたい。

### 3. 2. 質問紙データの結果と考察

段階反応モデルの識別力は部分得点モデルより全体的に高くなっている（表 7）。これと類似の現象はデータを 2 値化して通常の 2 母数ロジスティックモデルで分析した場合にも生じる。たとえば 4 段階評定の場合、2 値化は実質的には段階 2 を 1 に、段階 3 を 4 に置き換えるために生じる現象である。表 7 の「2 値モデル」はこのようなデータ変換をした結果である。部分得点モデルに比較してやや識別力が高くなっていることが分る。このような現象が発生することは、図 1 の項目 4 の群別の平均回答結果を 4 値と 2 値の場合で比較してみると良く理解できる。ただし図の作成にあたっては、4 値 2 値いずれの場合も全ての回答を 0 ~ 1 の範囲に変換している。多値データの 2 値化が識別力の推定値に大きな影響を与えることが図より理解できる。さて境界反応曲線が実質的には多値の回答を 2 値化して得られるデータに基づき定義され、推定されるため、この 2 値化と類似の現象が生じている可能性もある。しかしシミュレーションの結果では、部分得点モデルの識別力と段階反応モデルの識別力の推定値は大きく異なっていなかったのだから、2 値化を原因とすることにはやや疑問も残る。

もう一つの可能性として、テスト項目数の違いがある。シミュレーションでは 40 項目であったが、本質問紙は 9 項目であり、この違いが影響した可能性がある。このため 1 つシミュレーションを追加した。段階反応データのデータセット 1 の初めの 10 項目だけを取り、段階反応モデルにより交互同時推定値を求めたところ、識別力の値は表 8 のようになった。明らかに項目数が少なくなると識別力が大きくなる傾向があることが分る（ $MSE=0.2647$ ）。段階反応モデルでは少数項目での識別力の推定にバイアスがかかることは明らかとしても、その原因がモデルに存在するのか、それとも推定方法に問題があるのかは区別されなければならない。つまり段階反応モデルそのものではなくて、同時最尤推定という方法が推定に悪影響を及ぼしている可能性もある（同時最尤推定による項目母数の推定値が一致性を持たないことは良く知られている）。このため、MULTILOG（Thissen, 1991）により周辺最尤推定値も算出したが、バイアスはほとんど変わらない（ $MSE=0.2558$ ）。このことは推定法の問題というよりはモデルに問題があることを示唆するものであるかもしれない。なお表 8 からは項目

7 は一見すると同時最尤推定の識別力が極端に低く、反対に MULTILOG の推定値は妥当な値を得ているようであるが、実は MULTILOG は困難度において異常な推定値となっており、うまく推定できていないことには変わりはない。このため MSE の計算からは同項目を除外している。

- 5 部分得点データセット 1 の 10 項目を利用した部分得点モデルの分析では交互同時推定の MSE は、識別力が 0.0094、困難度が 0.0007 であり、周辺最尤推定では識別力が 0.0032、困難度が 0.0014 となった（表 9）。同モデルでは交互同時推定、周辺最尤推定ともこのバイアスは認められないとい
- 10 って良いだろう。表 10 には質問紙データを両モデルで分析し、周辺最尤推定を利用した推定値を示してある。同じく表 10 より、シミュレーションと同様に、周辺最尤推定を利用しても段階反応モデルでは識別力が大きくなっていることが分る。

- 段階反応モデルにおいて少数項目では識別力の上方バイアスがかかることについては、慎重な検討を要するので、これについては別の研究で行うこととし、こ
- 15 こでは表 7 の同時最尤推定の結果に基づいて検討を進める。図 2 に項目 1 の部分得点モデルの項目反応曲線を、また図 3 に段階反応モデルの反応曲線を示した。評定結果の合計点と両モデルの  $\theta$  のケンドールの順位相関を求めたところ段階反応モデルとは 0.9104、部分得点モデルとは 0.9424 の相関を得た（無回答の無いケース 1588 人のみで算出した。）。部分得点モデルの方が、合計
- 20 点との相関がやや高く、その再現には適していることはシミュレーションの結果と一致している。相関の値は段階反応データのシミュレーション結果から得られた両モデルの相関と近いものとなっているが、これだけで本データは段階反応モデルに従って発生していると決めるわけには行かないだろう。その理由としては、たとえばシミュレーションデータの作成の際に多次元性などの影響を加味すれば、
- 25 真値と推定値の相関だけでなく正答数（質問紙の場合は合計点）との相関も低下すると容易に予想できることなどがあげられるからである。

本節のデータは実際の調査であるため真値は不明であるからどちらのモデルが良いかを決定することは困難である。しかし段階反応モデルにおいて合計点との相関が低いことは、順位の逆転が十分説明がつくようなものであれば良いのであ

- るが、一見で分るような性格のものではないため、モデルの利用者には逆転現象を納得しにくいということは利用の際には考慮しておいた方がよい。いずれにしろシミュレーションの結果からは、（シミュレーションの条件に近い場合は）真の $\theta$ 値の推定という点では大きな差は両モデルでないことが想像される。しかし、
- 5 少数項目の場合の段階反応モデルの推定値のバイアスがどのような影響をもたらすのかは、もう少し検討する必要があるだろう。いずれにしろ項目特性として識別力や項目困難度以外に大きな関心がない場合、つまり回答の各段階の困難度などに特別な関心がない場合（質問紙を利用した多くの研究はこれに該当する）には、複雑で多くの母数を推定しなければならない段階反応モデルより簡単な部分
- 10 得点モデルの方が適していると言えるだろう。もちろん、個々の段階の反応にまで関心を持つ場合はこの限りではない。

#### 文献

- Andrich, D 1978 A rating formulation for ordered response categories. *Psychometrika*, 43, 561-573.
- 15 Birnbaum, A. 1968 Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores* (pp.395-479). Reading, MA: Addison-Wesley.
- 木原孝博・田中治彦・藤森進 1997 中学生・高校生の学校不適応に関する研究 岡山県青少年基本調査（1994）に基づく分析 岡山大学教育学部研究集録,
- 20 104, 105-122.
- 藤森進 2001 項目反応理論における部分得点の処理について 日本教育心理学会第43回総会発表論文集, 394.
- 藤森進 2002a 項目反応理論におけるテストの部分得点の処理方法について
- 25 未発表論文.
- 藤森進 2002b 部分得点モデルとその応用 第1回心理測定研究会.
- Masters, G.N. 1982 A Rasch model for partial credit scoring. *Psychometrika*, 47, 149-174.
- 野口裕之 1999 適応型テストへの応用：CAT方式による識別性検査（渡辺



直登・野口裕之編著「組織心理測定論」第8章 白桃書房)。

Samejima, F. 1969 Estimation of latent ability using a response pattern of graded scores. Psychometrika Monograph, No. 17.

- 5 Thissen, D 1991 Multilog user's guide. Chicago, IL: Scientific Software.

## 請求の範囲

1. 入力装置と出力装置とを備えた第1のコンピュータと、インターネットを含むネットワークを介して前記第1のコンピュータと接続されており入力装置と出力装置とを備えた第2のコンピュータと、前記ネットワークを介して前記第1及び第2のコンピュータと接続されたテスト管理サーバと、前記テスト管理サーバからアクセス可能であって難易度と識別力とを含む項目パラメータが予め推定されている複数の問題が記憶された問題データベースとによって構成され、1人の受験者に $n$ 問の問題を出題し、出題された $n$ 問の問題に対する前記受験者の反応から前記受験者の能力 $\theta$ を推定するテスト・システムであって、

前記テスト管理サーバは、

前記第1のコンピュータから送信されたリクエストに回答して、 $1 \leq j \leq n$ である問題 $j$ に対し満点を1として $0 \leq r_j \leq 1$ である部分得点 $r_j$ を許容する態様で採点されうる $n$ 問の問題を前記問題データベースから選択し前記第1のコンピュータに送信する手段と、

前記問題データベースから選択され前記第1のコンピュータに送信された問題に対して前記第1のコンピュータから返送された解答を記憶する解答記憶手段と、

前記第2のコンピュータから送信されたリクエストに回答して、前記解答記憶手段に記憶されている解答を読み出し前記第2のコンピュータに送信する手段と、

前記第2のコンピュータに送信された解答に対して与えられた部分得点 $r_j$ を前記第2のコンピュータから受信し記憶する部分得点記憶手段と、

前記部分得点記憶手段に記憶されている部分得点 $r_j$ と前記問題データベースに記憶されている問題 $j$ の項目パラメータとを用いて、当該部分得点 $r_j$ を獲得した受験者の能力 $\theta$ を推定する能力推定手段と、

を備えており、

前記能力推定手段において、 $P_j(\theta)$ は、部分得点 $r_j$ を問題 $j$ に固有であって前記受験者が正答1又は誤答0のいずれか一方の潜在的反応を取り得る潜在

的問題を  $s_j$  回繰り返し実施したときに前記受験者が潜在的に取りうる正誤反応の平均であると想定した場合に、前記受験者が前記潜在的問題に正答する確率であり、 $Q_j(\theta)$  は  $1 - P_j(\theta)$  である場合に、

【数 4 0】

$$5 \quad \ell_{part}(\theta) = \sum_{j=1}^n s_j (r_j \ln(P_j(\theta)) + (1 - r_j) \ln(Q_j(\theta)))$$

によって表される対数尤度  $\ell_{part}(\theta)$  を用いて受験者の能力  $\theta$  が推定されることを特徴とするテスト・システム。

2. 請求項 1 記載のテスト・システムにおいて、前記  $P_j(\theta)$  は、2 パラメータ・ロジスティック・モデルを用いて、

【数 4 1】

$$P_j(\theta) = \frac{1}{1 + \exp(-Da_j(\theta - b_j))}$$

と表現され、この数式 4 1 における  $a_j$  及び  $b_j$  はそれぞれが前記問題データベースに記憶されている問題が有する固有の特性である識別力と困難度とであり、  
15  $D$  は 1. 7 という定数であることを特徴とするテスト・システム。

3. 請求項 2 記載のテスト・システムにおいて、実際に観測される問題  $j$  に対する部分得点  $r_j$  が潜在的な複数の正誤問題の平均によって構成されている場合にも、これらの正誤問題に共通する正答確率を数式 4 1 で表現し、数式 4 0 を用いて受験者の能力  $\theta$  が推定されることを特徴とするテスト・システム。

4. 請求項 1 ないし請求項 3 の任意の請求項に記載のテスト・システムにおいて、テストが実施された集団の能力分布を仮定した上で、 $s_j$  回の正誤の和である 2 項分布と仮定された能力分布との積を能力の次元で積分して部分得点の理論的な分布関数を求め、求められた理論的な分布関数と実際のデータの部分得点の経験的な分布関数が最もよく一致するように、潜在的な問題の繰り返し回数である  $s_j$  を推定する手段を更に備えていることを特徴とするテストシステム。

5. 請求項 1 ないし請求項 4 の任意の請求項に記載のテスト・システムにおいて、前記第 1 及び第 2 のコンピュータにおける前記出力装置と前記入力装置とはそれぞれ音声出力装置と音声入力装置とを含み、前記テストサーバに送信され記憶される解答は音声データを含むことを特徴とするテスト・システム。

6. 入力装置と出力装置とを備えた第 1 のコンピュータと、インターネットを含むネットワークを介して前記第 1 のコンピュータと接続されており入力装置と出力装置とを備えた第 2 のコンピュータと、前記ネットワークを介して前記第 1 及び第 2 のコンピュータと接続されたテスト管理サーバと、前記テスト管理サーバからアクセス可能であって難易度と識別力とを含む項目パラメータが予め推定されている複数の問題が記憶された問題データベースとによって構成され、1 人の受験者に  $n$  問の問題を出題し、出題された  $n$  問の問題に対する前記受験者の反応から前記受験者の能力  $\theta$  を推定するテスト・システムを制御する方法であって、前記テスト管理サーバにおいて、

(1) 前記第 1 のコンピュータから送信されたリクエストに回答して、 $1 \leq j \leq n$  である問題  $j$  に対し満点を 1 として  $0 \leq r_j \leq 1$  である部分得点  $r_j$  を許容する態様で採点されうる  $n$  問の問題を前記問題データベースから選択し前記第 1 のコンピュータに送信するステップと、

(2) 前記問題データベースから選択され前記第 1 のコンピュータに送信された問題に対して前記第 1 のコンピュータから返送された解答を記憶するステップと、

(3) 前記第 2 のコンピュータから送信されたリクエストに回答して、前記ステップ (2) において記憶された解答を読み出し前記第 2 のコンピュータに送信する手段と、

(4) 前記第 2 のコンピュータに送信された解答に対して与えられた部分得点  $r_j$  を前記第 2 のコンピュータから受信し記憶するステップと、

(5) 前記ステップ (4) において記憶された部分得点  $r_j$  と前記問題データベースに記憶されている問題  $j$  の項目パラメータとを用いて、当該部分得点  $r$

、 $j$  を獲得した受験者の能力  $\theta$  を推定するステップと、  
を含んでおり、

前記ステップ (5) において、 $P_j(\theta)$  は、部分得点  $r_j$  を問題  $j$  に固有であって前記受験者が正答 1 又は誤答 0 のいずれか一方の潜在的反応を取り得る潜在的問題を  $s_j$  回繰返し実施したときに前記受験者が潜在的に取りうる正誤反応の平均であると想定した場合に、前記受験者が前記潜在的問題に正答する確率であり、 $Q_j(\theta)$  は  $1 - P_j(\theta)$  である場合に、

【数 4 2】

$$\ell_{part}(\theta) = \sum_{j=1}^n s_j (r_j \ln(P_j(\theta)) + (1 - r_j) \ln(Q_j(\theta)))$$

10 によって表される対数尤度  $\ell_{part}(\theta)$  を用いて受験者の能力  $\theta$  が推定されることを特徴とするテスト・システム制御方法。

7. 請求項 6 記載の方法において、前記  $P_j(\theta)$  は、2 パラメータ・ロジスティック・モデルを用いて、

15 【数 4 3】

$$P_j(\theta) = \frac{1}{1 + \exp(-Da_j(\theta - b_j))}$$

と表現され、この数式 4 3 における  $a_j$  及び  $b_j$  はそれぞれが前記問題データベースに記憶されている問題が有する固有の特性である識別力と困難度とであり、 $D$  は 1.7 という定数であることを特徴とする方法。

20

8. 請求項 7 記載の方法において、実際に観測される問題  $j$  に対する部分得点  $r_j$  が顕在的な複数の正誤問題の平均によって構成されている場合にも、これらの正誤問題に共通する正答確率を数式 4 3 で表現し、数式 4 2 を用いて受験者の能力  $\theta$  が推定されることを特徴とする方法。

25

9. 請求項 6 ないし請求項 8 記載の方法において、

(6) テストが実施された集団の能力分布を仮定した上で、 $s_j$  回の正誤の和で

ある 2 項分布と仮定された能力分布との積を能力の次元で積分して部分得点の理論的な分布関数を求め、求められた理論的な分布関数と実際のデータの部分得点の経験的な分布関数が最もよく一致するように、潜在的な問題の繰り返し回数である  $J$

図 1

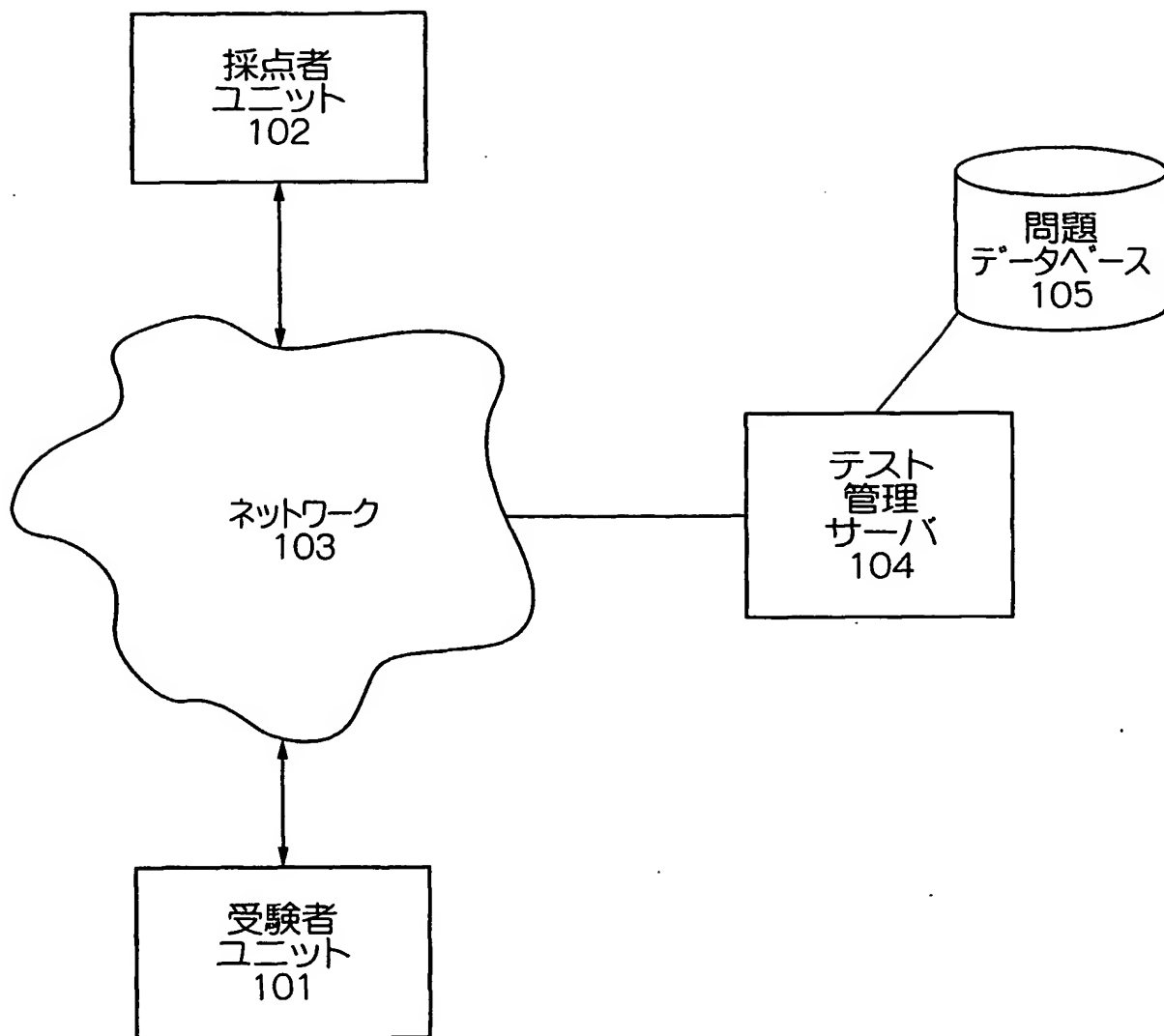


図 2

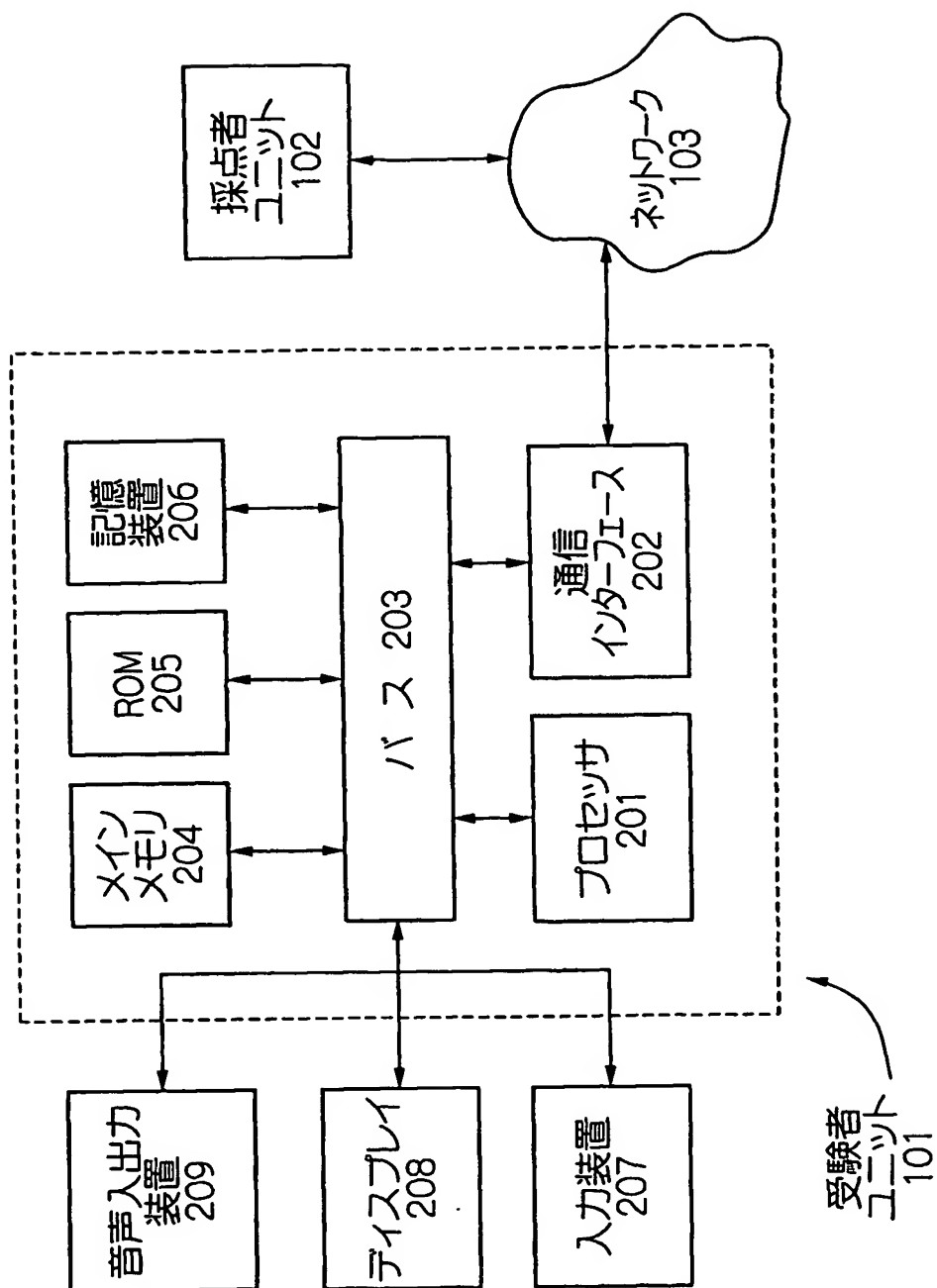




図 3

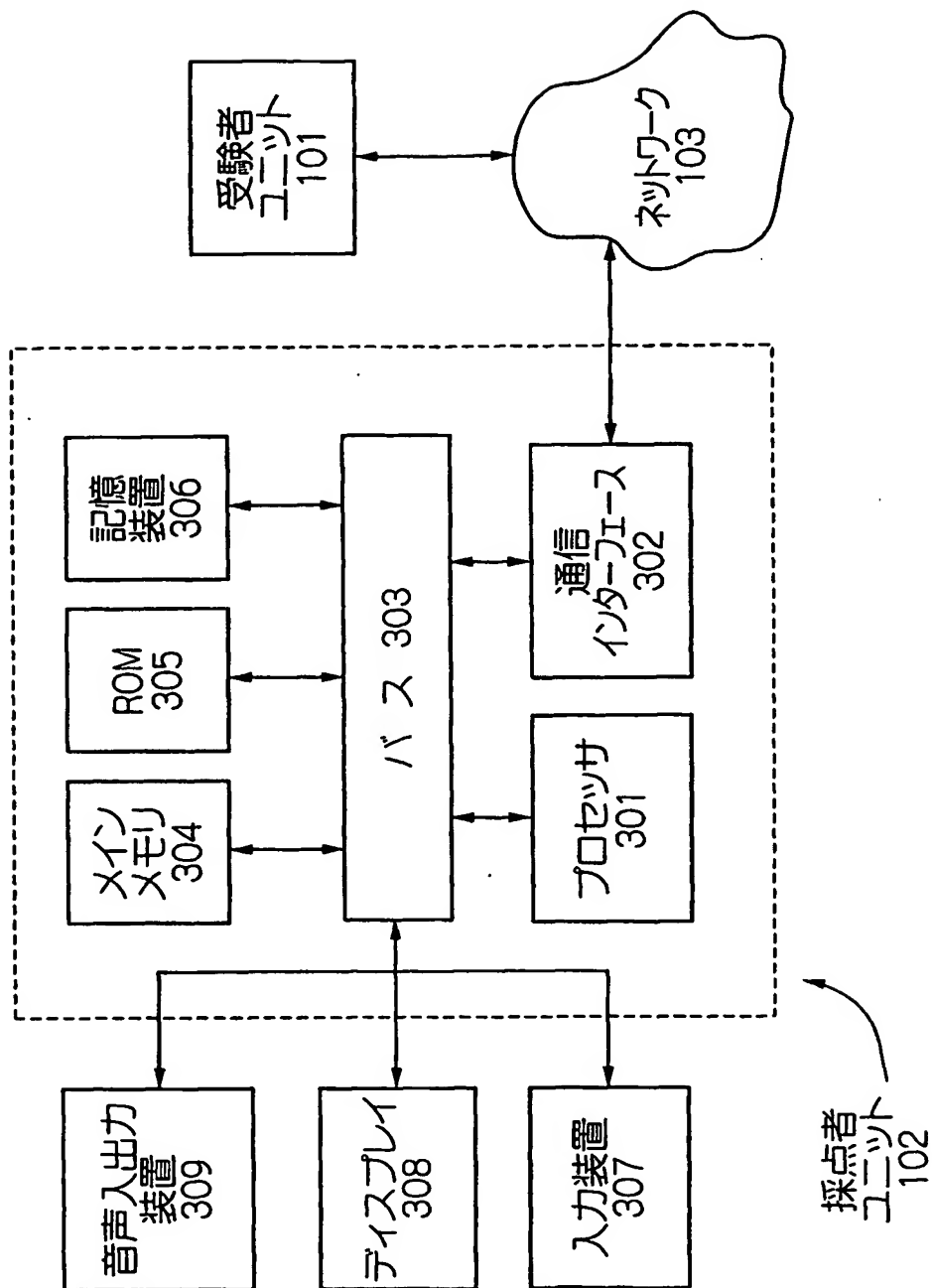


図4

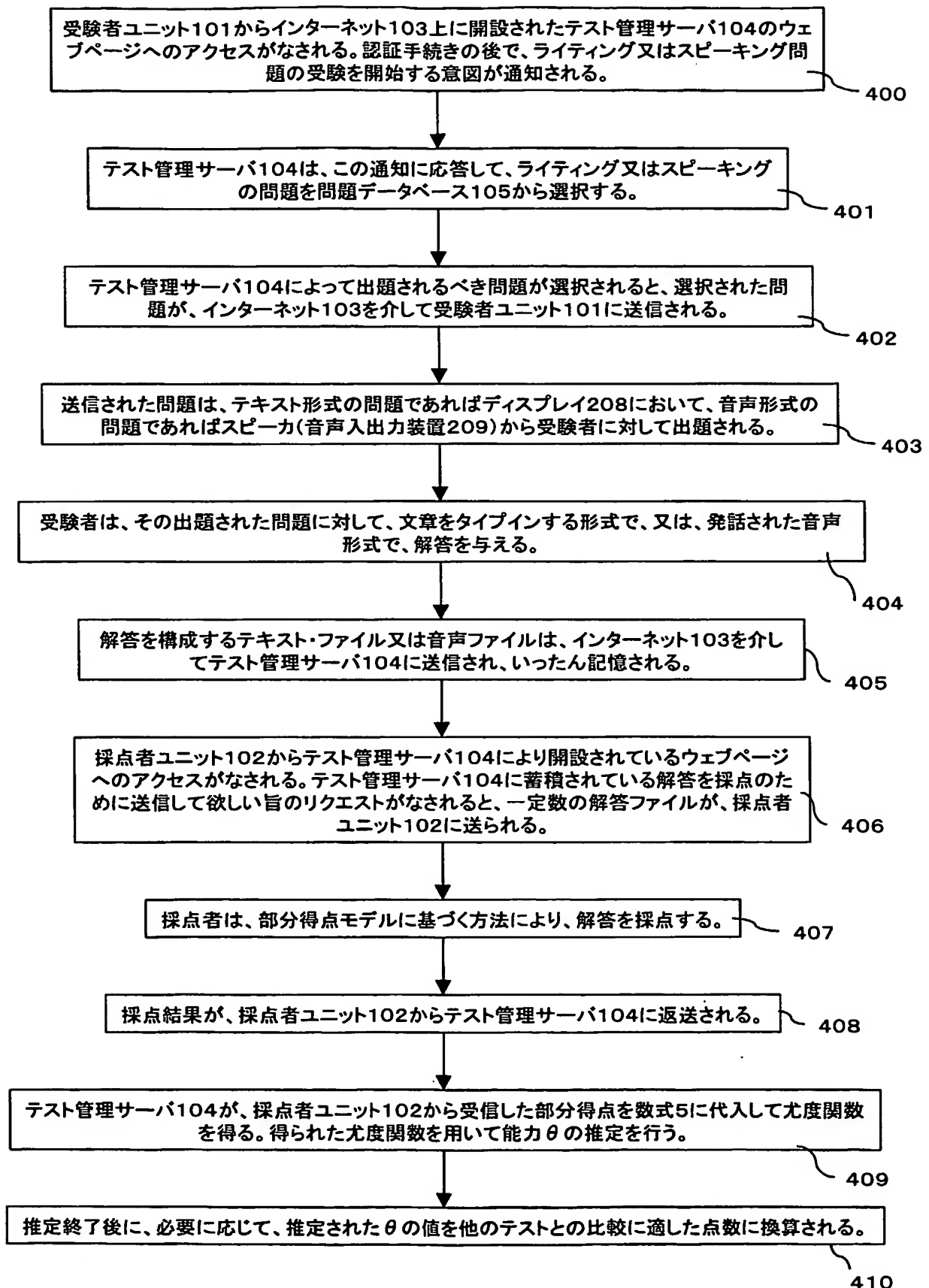


図 5

	1 回目平均	2 回目平均	3 回目平均	変動幅
リスニング (250 点満点)	154. 2	160. 8	158. 2	6. 6
リーディング (250 点満点)	158. 5	166. 9	177. 0	18. 5
ライティング (250 点満点)	147. 1	150. 9	154. 8	7. 7
スピーキング (250 点満点)	142. 1	147. 5	142. 3	5. 4
トータルスコア (1000 点満点)	601. 8	625. 9	632. 3	30. 5

図 6 a

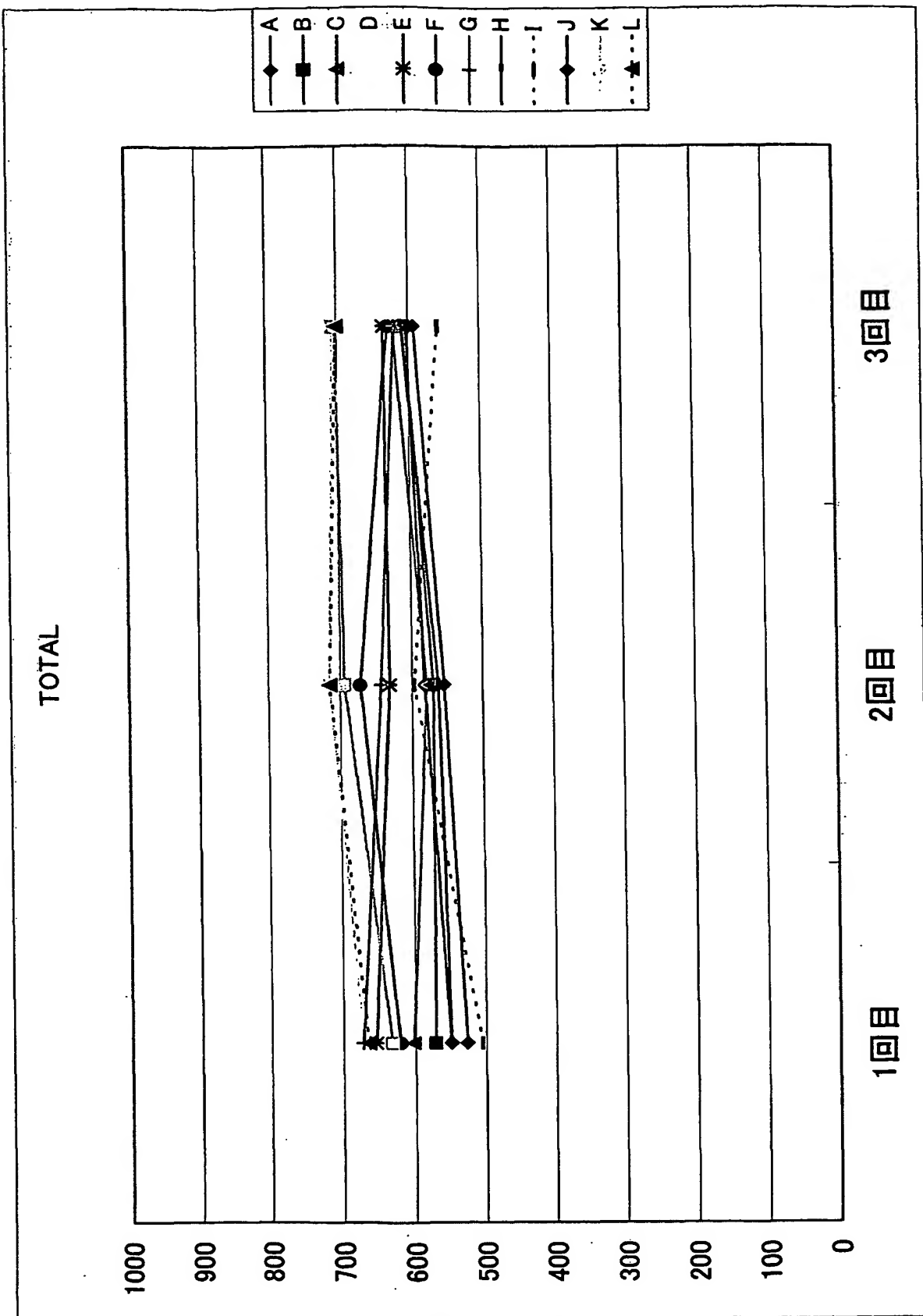


図 6 b

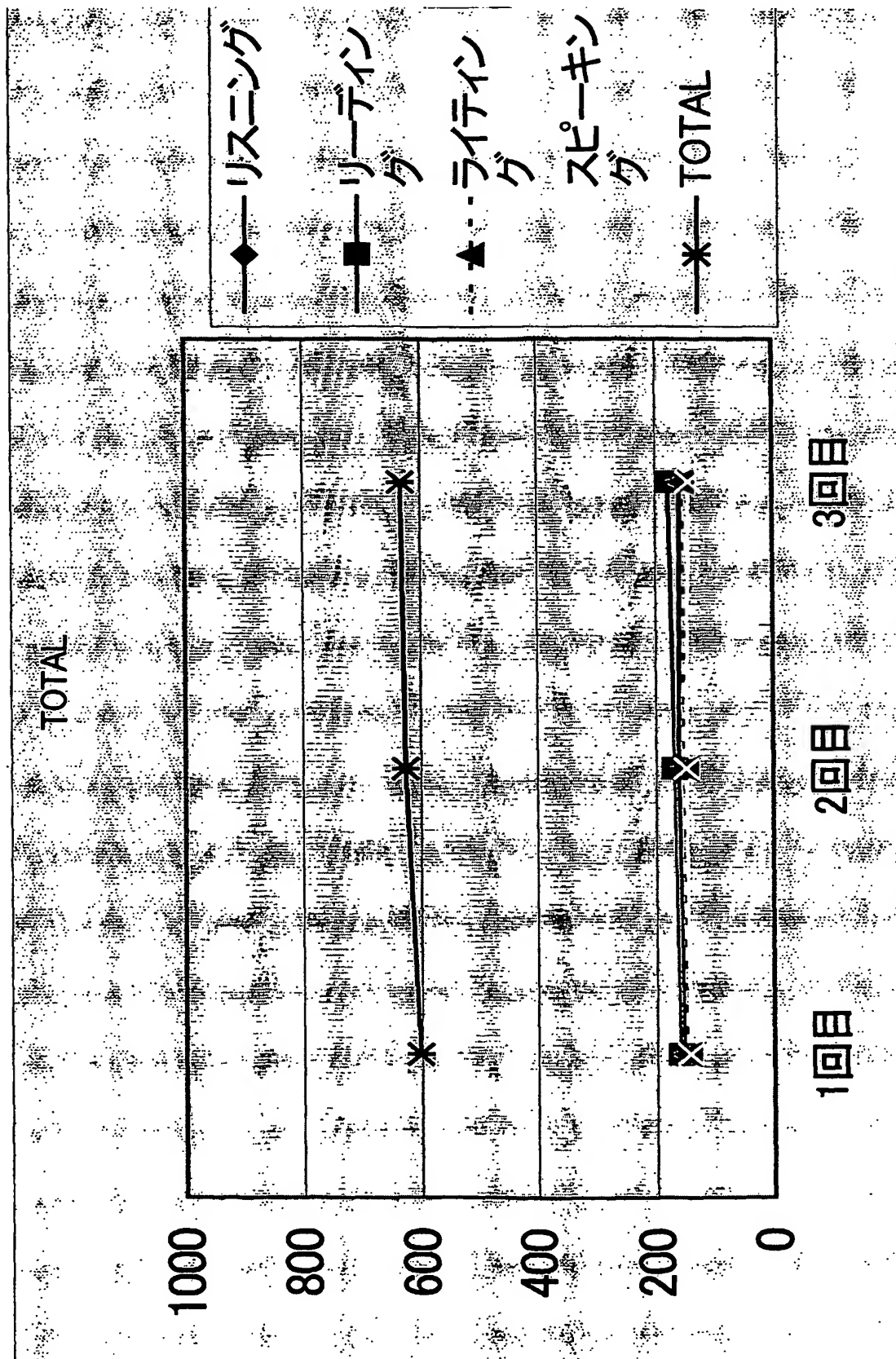


図 6 c

## 最高点と最低点の差

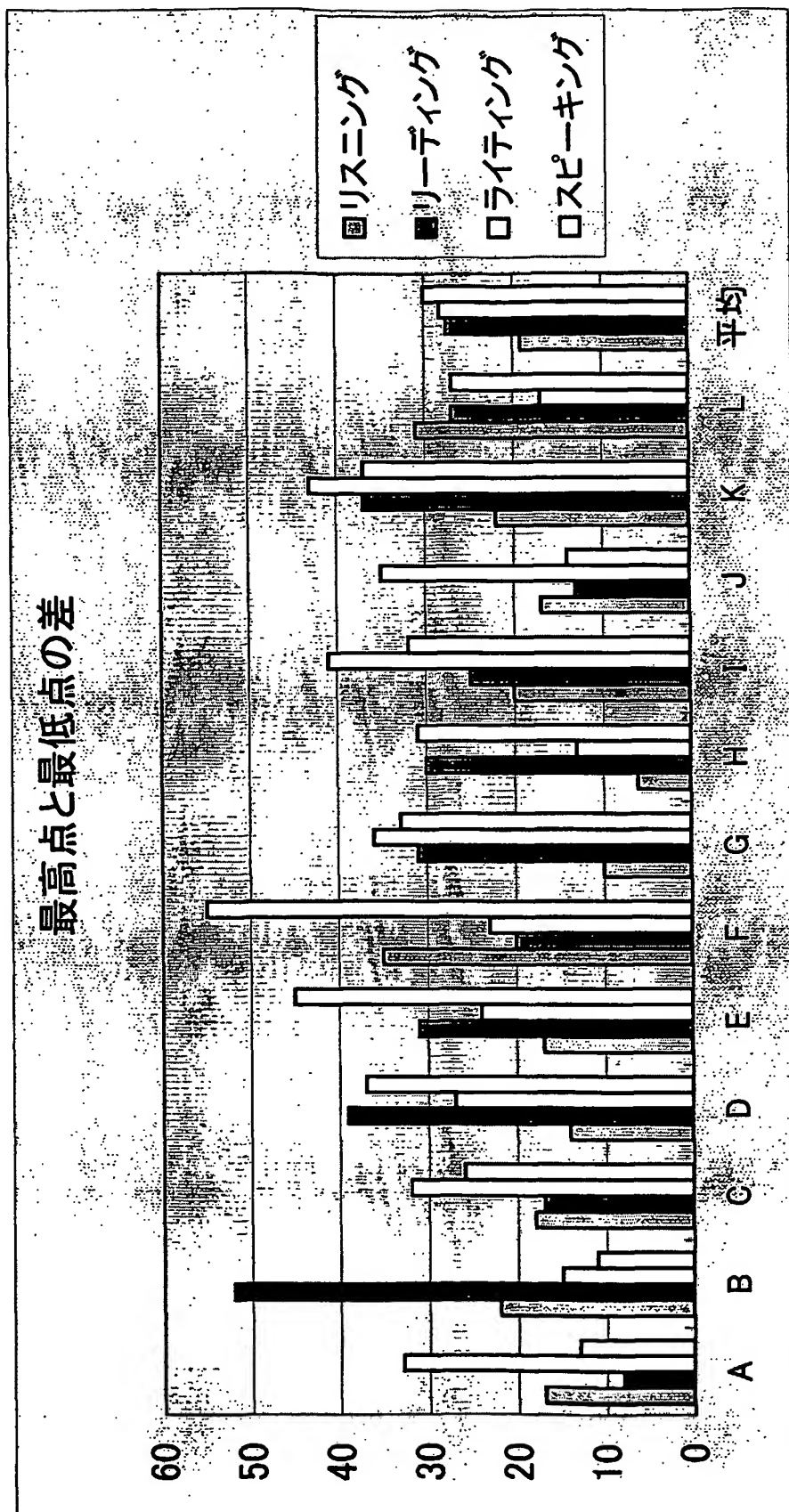


図6d

リスニング

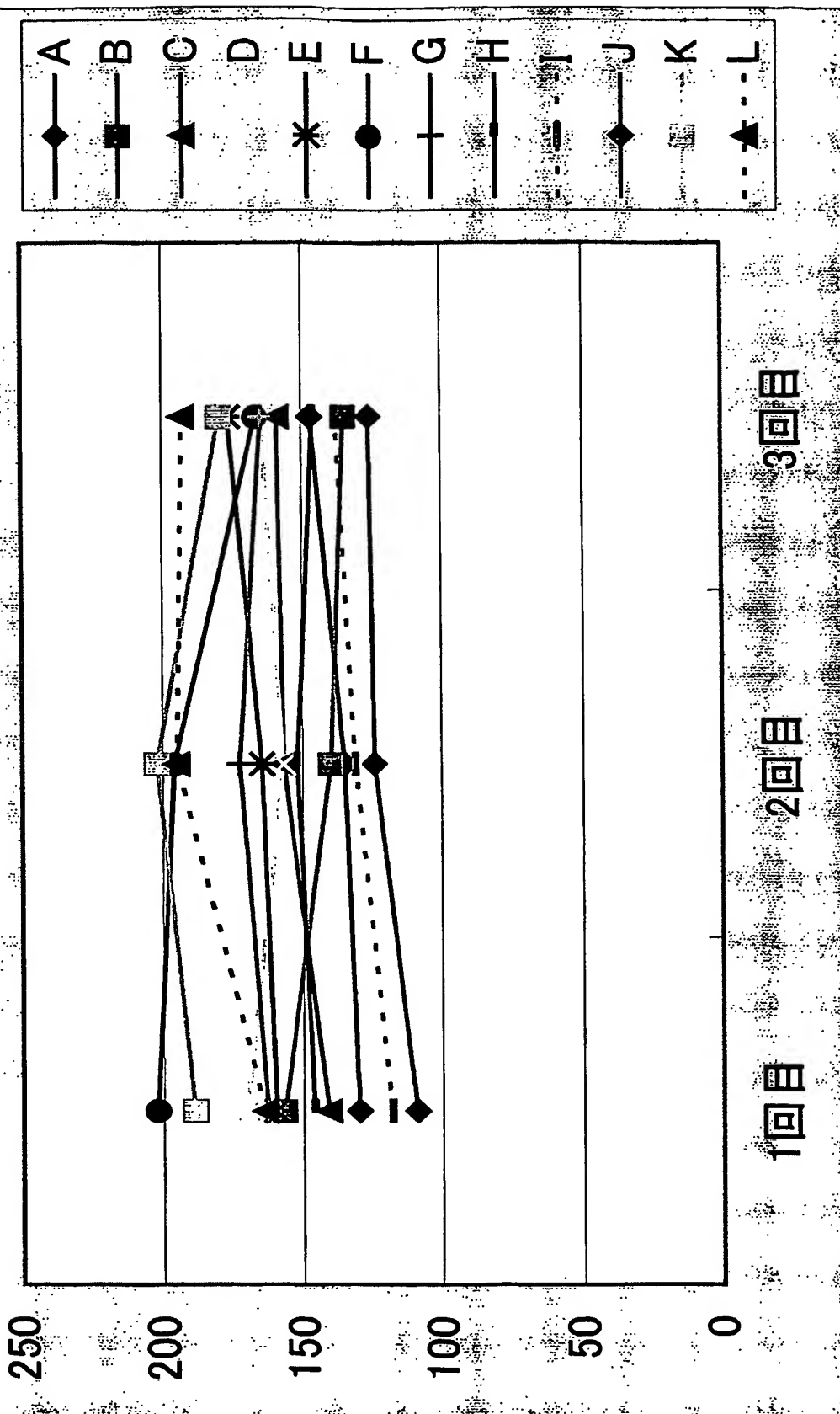


図 6 e

リーディング

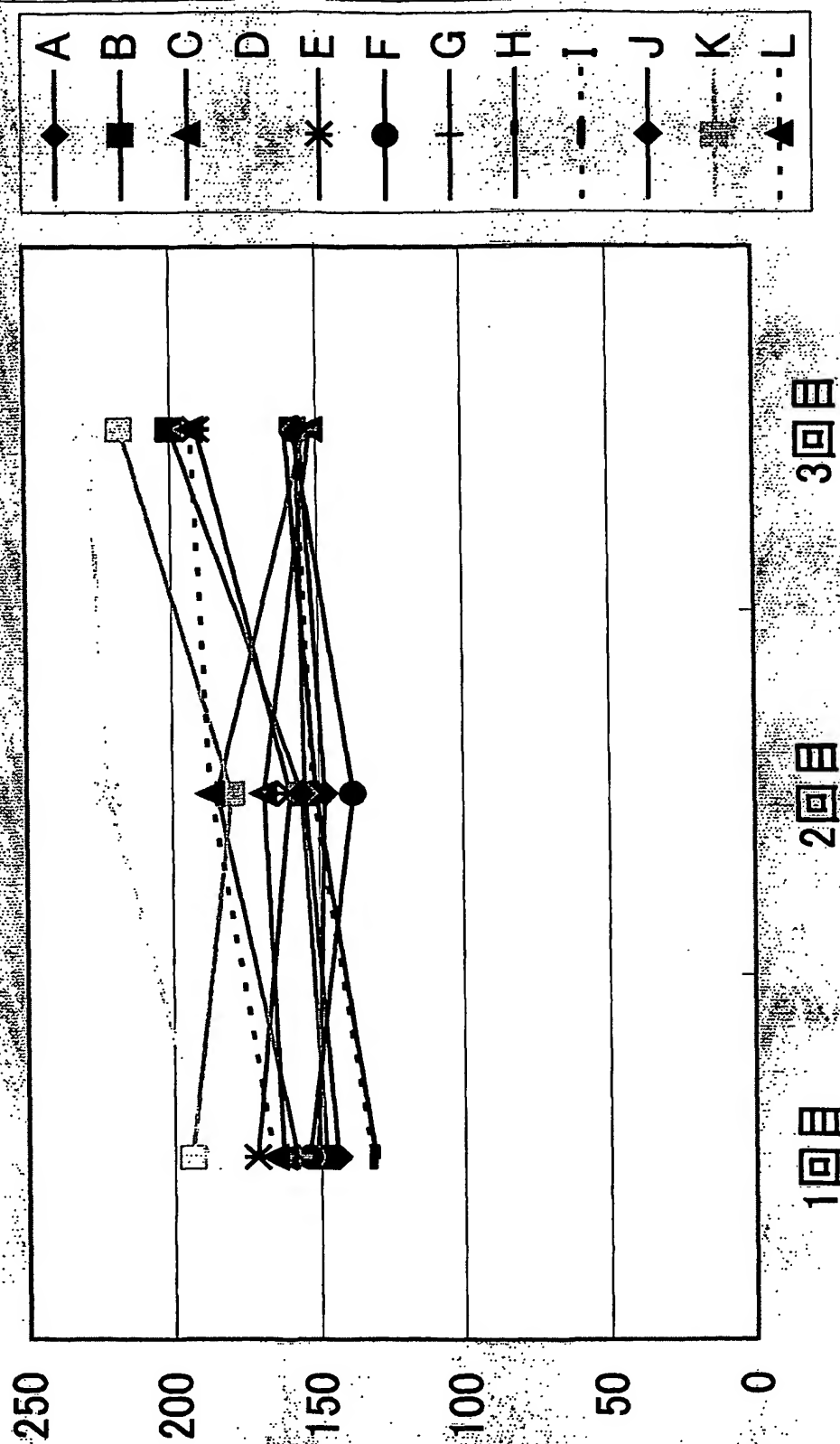




図6 f

ライティング

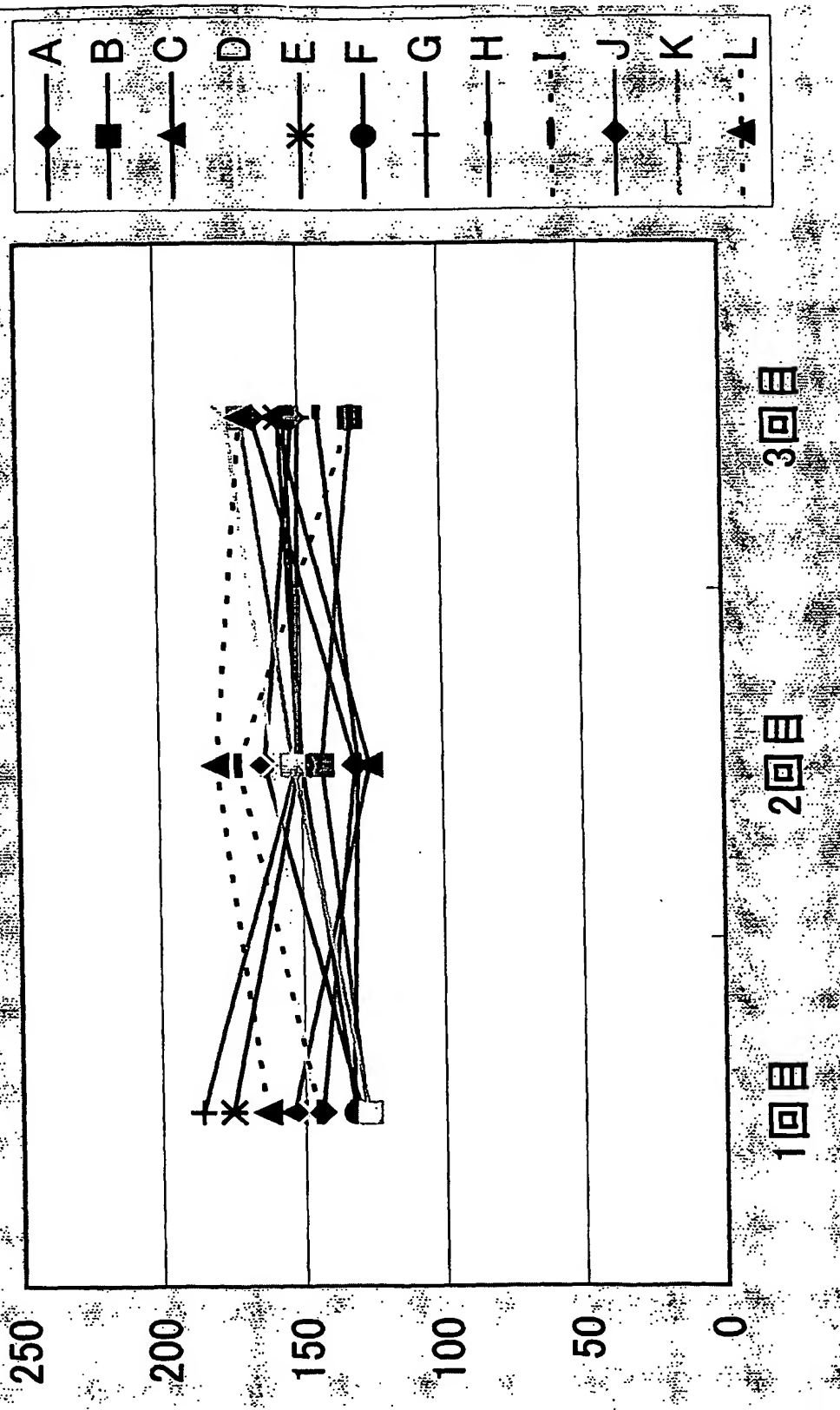
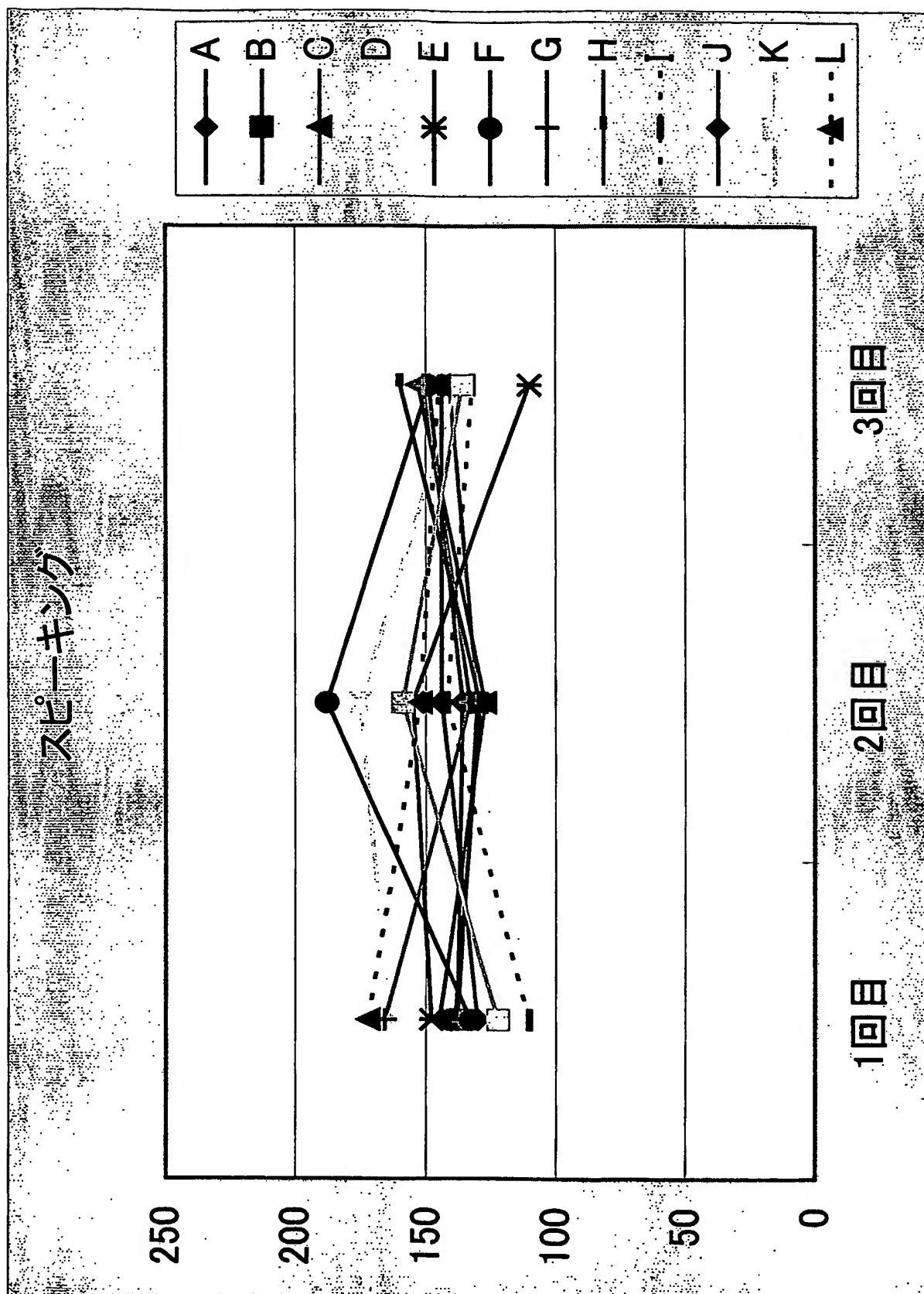


図 6 g



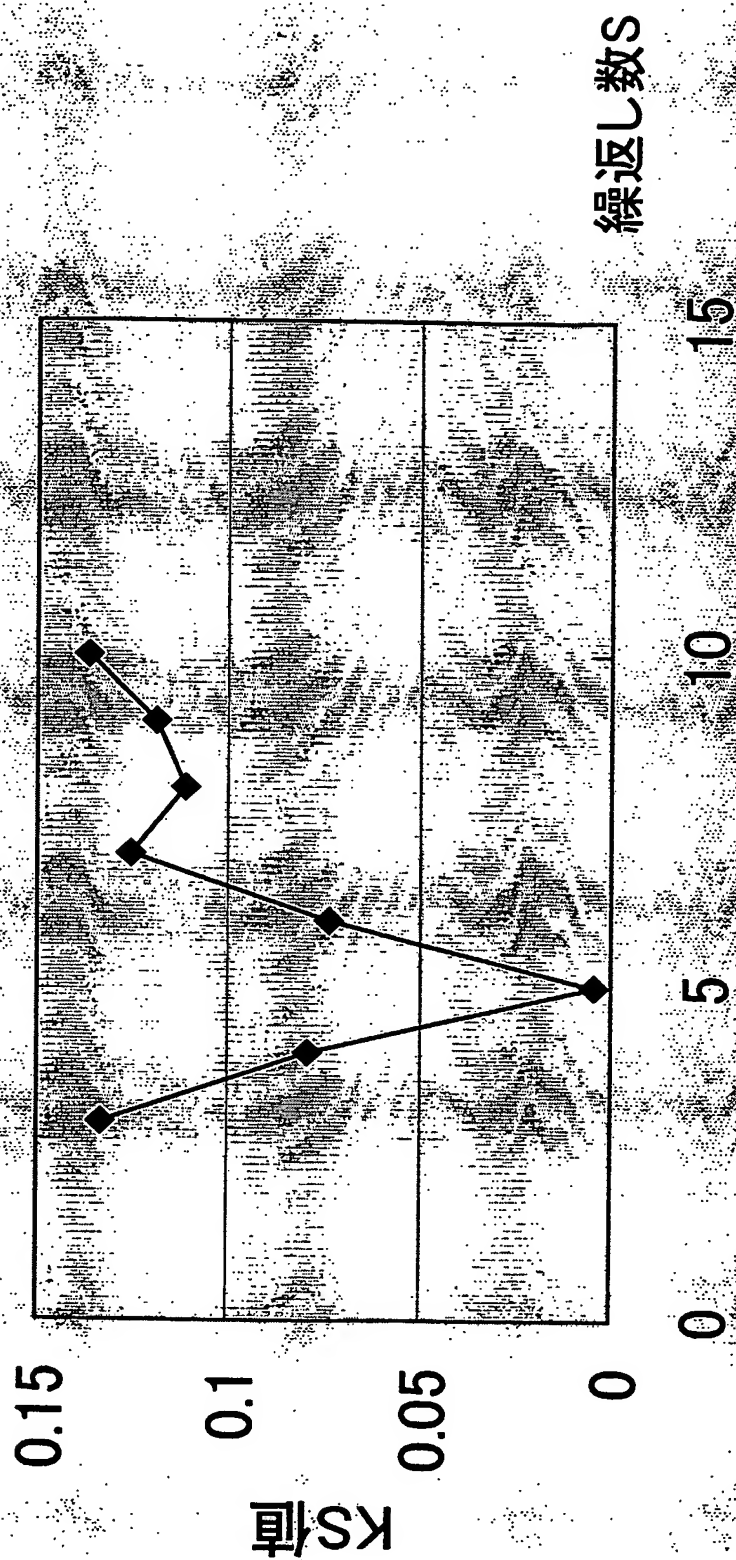
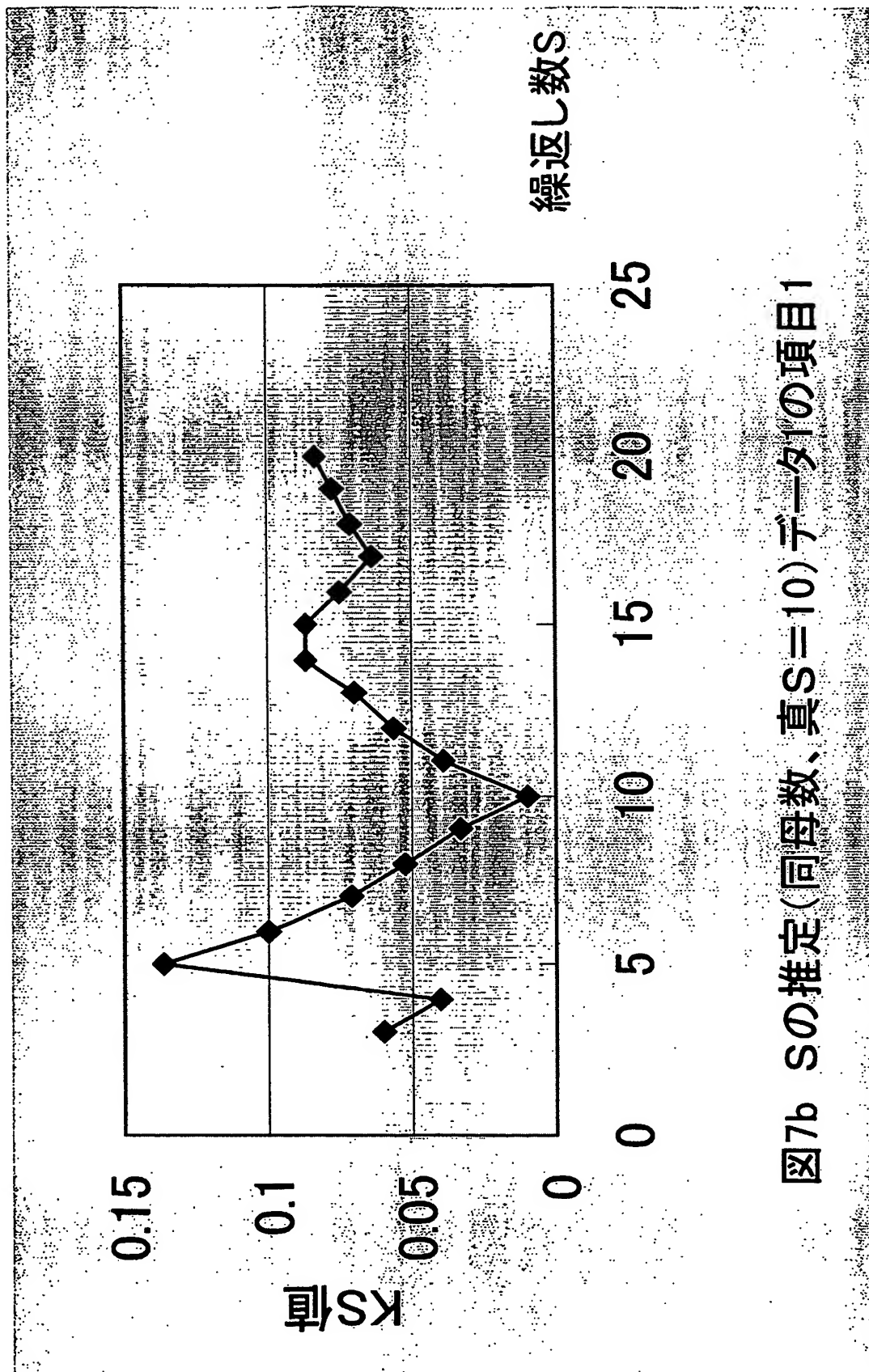


図7a Sの推定(同母数、真 $S=5$ )データ1の項目1

図7b Sの推定(同母数、真 $S=10$ )データ1の項目1

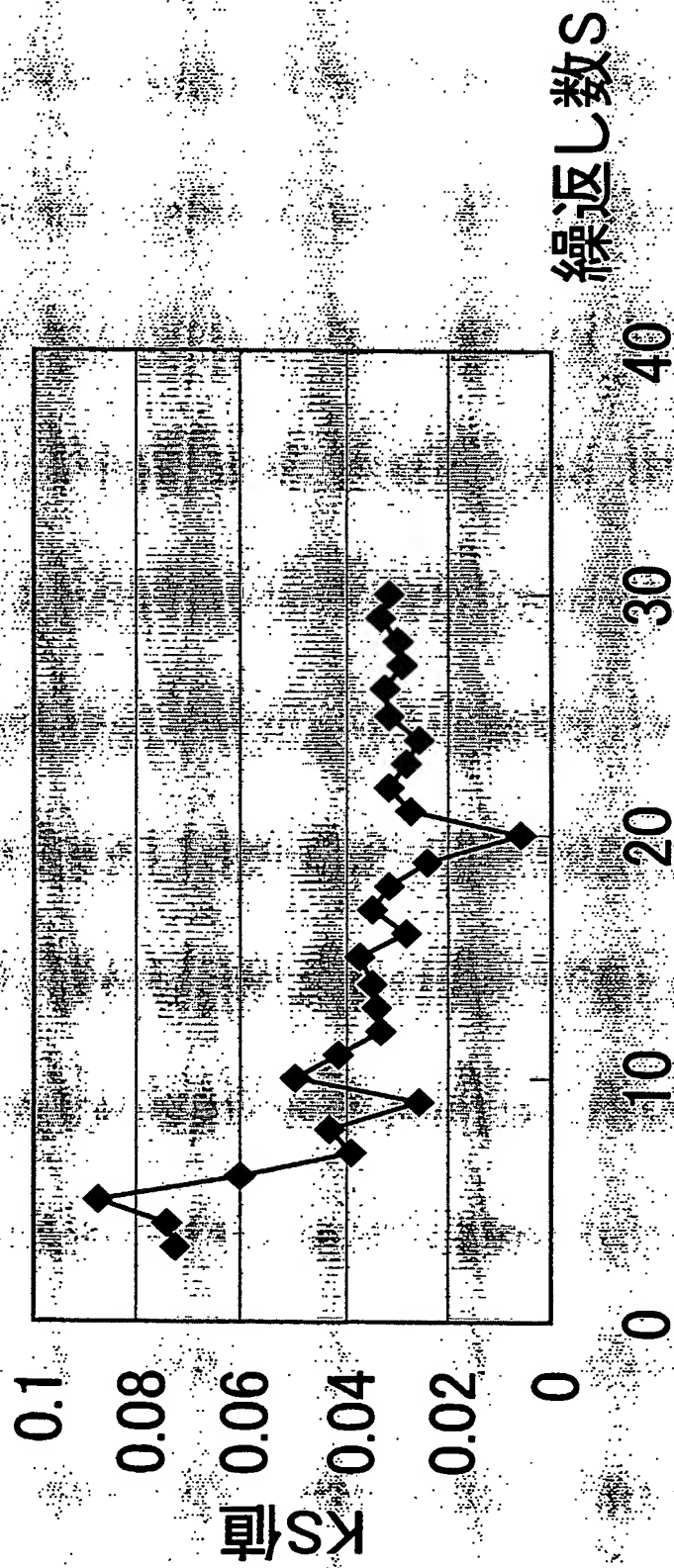


図7c Sの推定(同母数、真 $S=20$ )データ1の項目1

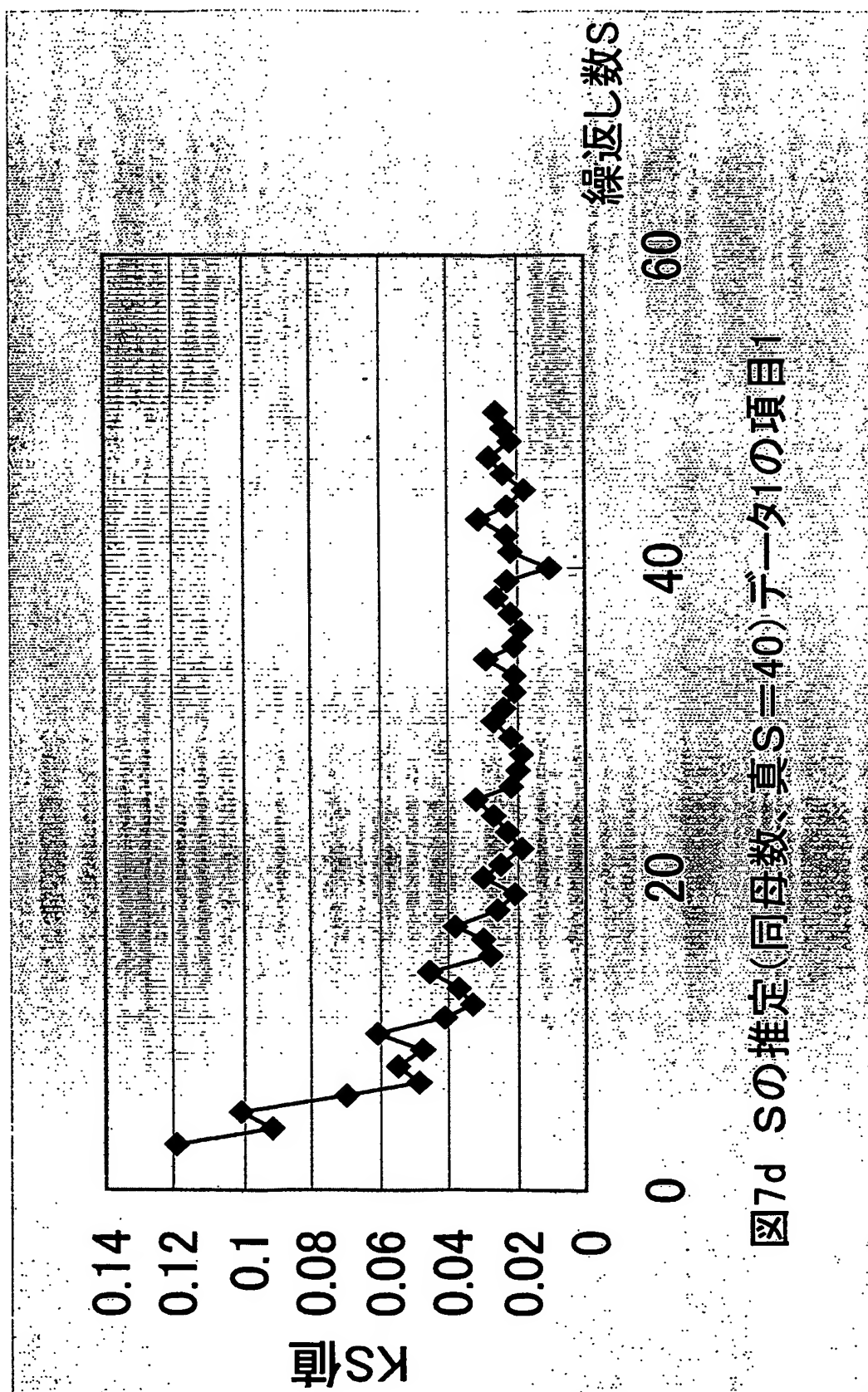


図7d Sの推定(同母数、真S=40)データ1の項目1

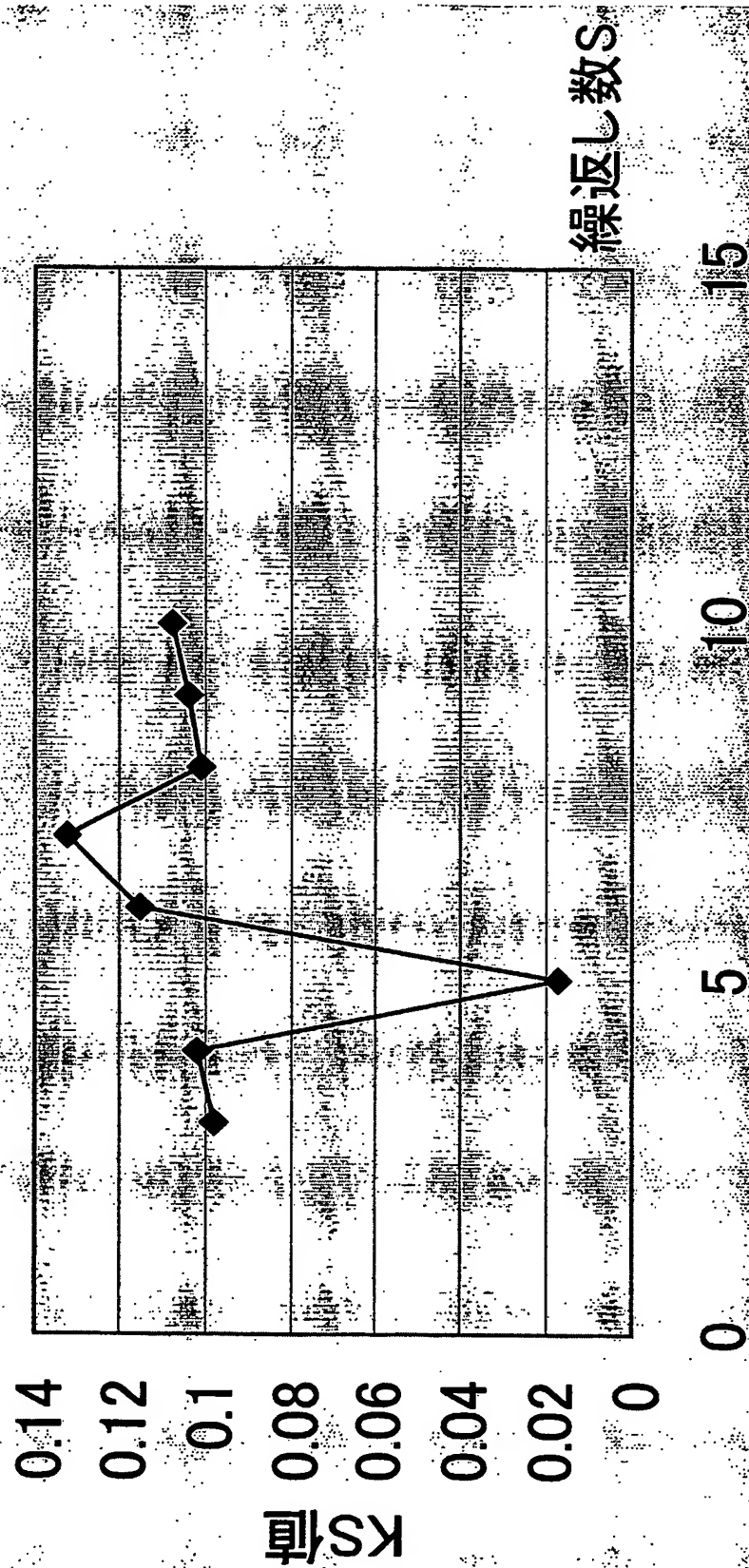


図8a Sの推定(B社英語テスト部分得点項目31)



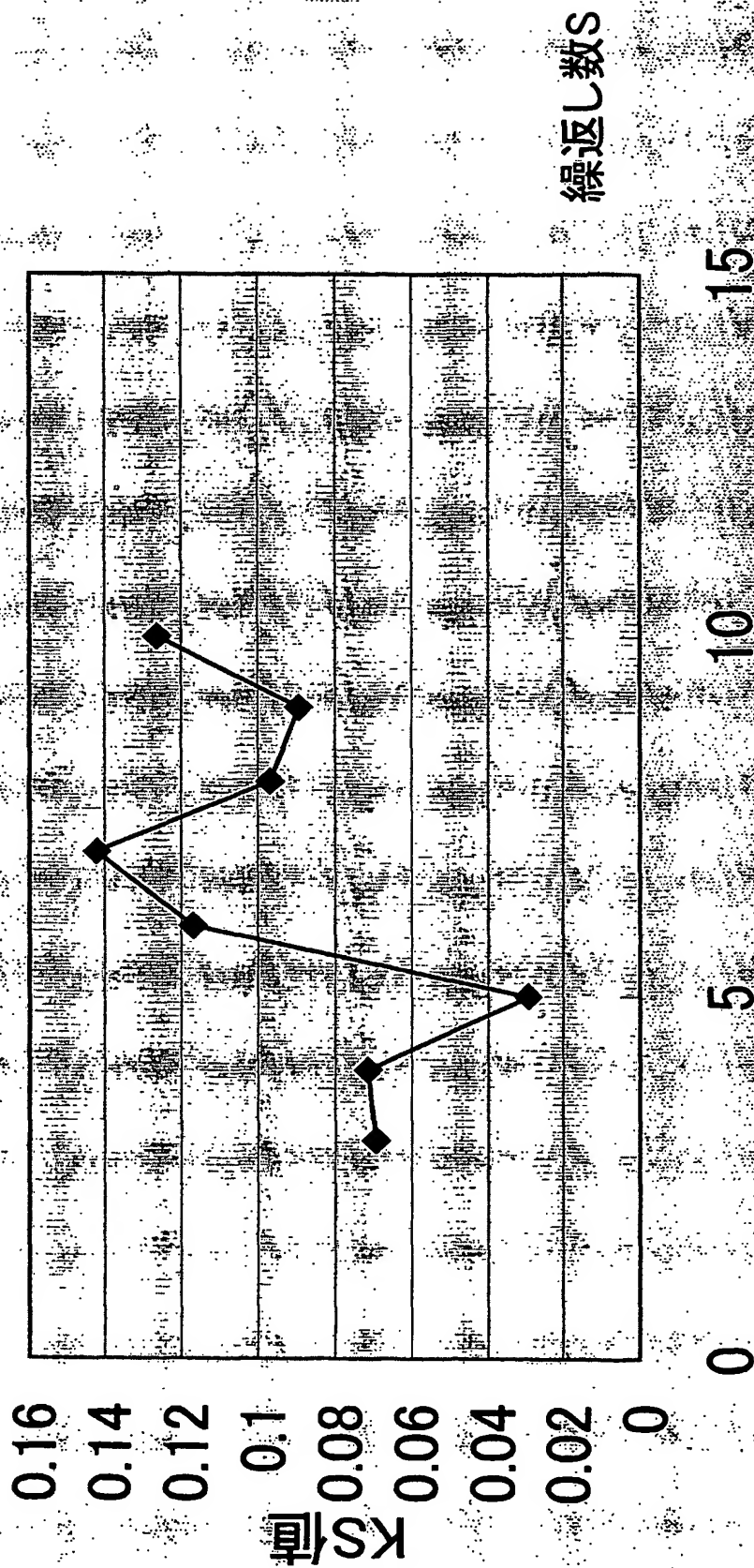


図8b Sの推定(B社英語テスト部分得点項目32)



# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/JP03/12252

<b>A. CLASSIFICATION OF SUBJECT MATTER</b> Int.Cl <sup>7</sup> G09B19/00, G06F17/18, G06F17/60		
According to International Patent Classification (IPC) or to both national classification and IPC		
<b>B. FIELDS SEARCHED</b> Minimum documentation searched (classification system followed by classification symbols) Int.Cl <sup>7</sup> G09B5/00-7/12, G09B19/00, G09B19/06, G06F17/10, G06F17/18, G06F17/60		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Jitsuyo Shinan Koho 1922-1996 Toroku Jitsuyo Shinan Koho 1994-2003 Kokai Jitsuyo Shinan Koho 1971-2003 Jitsuyo Shinan Toroku Koho 1996-2003		
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)		
<b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2002-6734 A (The Society for Testing English Proficiency, Inc.), 11 January, 2002 (11.01.02), Full text; all drawings (Family: none)	1-9
A	Giichiro SUZUKI "Graphic Tokeigaku", 1st edition, Jikkyo Shuppan Co., Ltd., 28 November, 1997 (28.11.97), ISBN 4-407-02390-2, pages 122 to 124	1-9
A	JP 2001-56634 A (Toshiba Corp.), 27 February, 2001 (27.02.01), Full text; all drawings (Family: none)	1-9
<input type="checkbox"/> Further documents are listed in the continuation of Box C. <input type="checkbox"/> See patent family annex.		
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier document but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed		"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family
Date of the actual completion of the international search 19 December, 2003 (19.12.03)		Date of mailing of the international search report 20 January, 2004 (20.01.04)
Name and mailing address of the ISA/ Japanese Patent Office		Authorized officer
Facsimile No.		Telephone No.

## A. 発明の属する分野の分類 (国際特許分類 (IPC))

Int. Cl<sup>7</sup> G09B 19/00, G06F 17/18, G06F 17/60

## B. 調査を行った分野

調査を行った最小限資料 (国際特許分類 (IPC))

Int. Cl<sup>7</sup> G09B 5/00 - 7/12, G09B 19/00, G09B 19/06,  
G06F 17/10, G06F 17/18, G06F 17/60

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報	1922-1996年
日本国公開実用新案公報	1971-2003年
日本国登録実用新案公報	1994-2003年
日本国実用新案登録公報	1996-2003年

国際調査で使用した電子データベース (データベースの名称、調査に使用した用語)

## C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
A	JP 2002-6734 A (財団法人日本英語検定協会) 2002.01.11 全文, 全図 (ファミリーなし)	1-9
A	鈴木 義一郎, グラフィック統計学 初版, 実教出版株式会社, 1997.11.28, ISBN 4-407-02390-2, p.122~124	1-9
A	JP 2001-56634 A (株式会社東芝) 2001.02.27 全文, 全図 (ファミリーなし)	1-9

☐ C欄の続きにも文献が列挙されている。☐ パテントファミリーに関する別紙を参照。

## \* 引用文献のカテゴリー

「A」 特に関連のある文献ではなく、一般的技術水準を示すもの  
「E」 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの  
「L」 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献 (理由を付す)  
「O」 口頭による開示、使用、展示等に言及する文献  
「P」 国際出願日前で、かつ優先権の主張の基礎となる出願

の日の後に公表された文献  
「T」 国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの  
「X」 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの  
「Y」 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの  
「&」 同一パテントファミリー文献

国際調査を完了した日

19.12.03

国際調査報告の発送日

20.01.04

国際調査機関の名称及びあて先

日本国特許庁 (ISA/JP)  
郵便番号 100-8915  
東京都千代田区般が関三丁目4番3号

特許庁審査官 (権限のある職員)  
官本 昭彦



2T 9226

電話番号 03-3581-1101 内線 3277

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☒ FADED TEXT OR DRAWING
- ☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☒ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**